MICROBIOLOGICAL REVIEWS, June 1993, p. 320–346
0146-0749/93/020320-27$02.00/0

Vol. 57, No. 2

Structural, Functional, and Evolutionary Relationships among Extracellular Solute-Binding Receptors of Bacteria

ROLAND TAM AND MILTON H. SAIER, JR.*

Department of Biology, University of California at San Diego, La Jolla, California 92093-0116

INTRODUCTION

"Everything should be made as simple as possible, but not simpler."

A. Einstein

Periplasmic solute-binding proteins of gram-negative bacteria and homologous, external membrane-bound lipoproteins of gram-positive bacteria are generally known to participate in two related recognition phenomena: chemoreception and transmembrane transport (46, 55, 63, 78, 191). In cases of both transport and chemotaxis, the hydrophilic solute-binding domains of these proteins interact with integral cytoplasmic membrane proteins following a conformational change induced by solute binding. Solute binding is known to induce a substantial conformational change in the periplasmic binding proteins of gram-negative bacteria, which facilitates interaction with the transmembrane chemoreception or transport protein(s) (120, 129, 132, 133, 154). The integral membrane chemoreception proteins, termed methylated chemoreception proteins, span the membrane twice (97, 191), whereas the integral membrane constituents of the periplasmic-binding-protein-dependent transport systems, of which there are two per permease, usually span the membrane five or six times (96, 159, 171). Limited sequence comparison studies, as well as comparisons of three-dimensional structures of the periplasmic binding proteins, have led to the suggestion that at least some of these proteins share a evolutionary origin (3, 47, 131, 154, 187, 209). Members of one class of periplasmic binding proteins also appear to be related in sequence, structure, and evolutionary origin to a class of bacterial DNA-binding proteins which function in transcriptional regulation (122, 123, 207, 212).

Recent evidence suggests that in addition to their roles in transport and chemoreception, at least some solute-binding proteins function in the initiation of sensory transduction pathways. They apparently detect signals in the external environment and transmit these signals via their cognate transmembrane transport proteins or other transmembrane proteins to cytoplasmic constituents. Cytoplasmic proteins then signal a change in state, usually by altering the rates of specific gene expression. A periplasmic phosphate-binding protein of enteric bacteria apparently initiates the signal which results in induction of the phosphate (*pho*) regulon. In agrobacterial species, extracellular signaling between plant hosts and the bacterial parasite is dependent on periplasmic binding proteins. In *Salmonella typhimurium*, virulence is apparently dependent on a binding-protein-dependent transporter which is homologous to the well-characterized bacterial peptide transport systems. Sporulation initiation in *Ba-*

* Corresponding author.

cillus species appears to be regulated by two distinct peptide transporters, both of which depend on solute-binding proteins. Although these two transport systems are similar in structure, the physiological consequences of their actions are opposing. It is therefore clear that external solute-binding proteins perform receptor functions not only in transport and chemoreception but also in signal transduction. In this review the recent advances in the role of binding proteins in sensory transduction will be briefly discussed. The reader is referred to other comprehensive reviews for detailed discussions of the roles of these proteins in transport and chemotaxis (9, 21, 55, 59, 75, 113, 191).

Although limited sequence analyses of specific binding proteins have been reported, as discussed in the preceding paragraph as well as in several of the references listed in Table 1, no comprehensive study of these proteins, aimed at their systematic classification, has been reported. The present study was designed to correct this deficiency. Over 50 sequenced periplasmic binding proteins of gram-negative bacteria and homologous lipoproteins of gram-positive bacteria were analyzed by standard computer methods, and their degrees of relatedness were defined. The comparison scores reported establish that some of these proteins are clearly related by descent, that others are very probably related by descent, and that others exhibit too great a degree of sequence divergence to allow establishment of a common evolutionary origin on the basis of sequence similarities alone. The vast majority of these solute-binding proteins can be grouped into eight families or clusters, which generally correlate with the molecular size of the protein and the nature of the solute bound. Cluster 1 proteins bind oligosaccharides, glycerol 3-phosphate, and iron. Cluster 2 proteins are specific for pentoses and hexoses and are homologous to a family of bacterial transcription factors, most of which are known to bind sugars (122, 123, 207, 212). Cluster 3 proteins recognize polar amino acids and opines. Cluster 4 proteins bind aliphatic, hydrophobic amino acids and are homologous to a bacterial transcription factor, the AmiC repressor of aliphatic amidase expression in *Pseudomonas aeruginosa* (219). Cluster 5 proteins are specific for peptides and nickel. Cluster 6 consists of multivalent-inorganic-anion-binding proteins, whereas cluster 7 includes two divergent, multivalent-organic-polyanion-binding proteins. Finally, cluster 8 proteins exhibit specificity for organic iron complexes and may include the vitamin B_{12}-binding protein. The members of each of these clusters of binding proteins exhibit maximal sequence conservation in their N-terminal domains, and signature sequences, characteristic of the individual families of binding proteins, have been derived from these conserved regions. Only 1 of the more than 50 proteins analyzed exhibits demonstrable homology to any of the thousands of sequenced eukaryotic proteins in the current data bases, and this protein, the vitamin B_{12}-binding protein of *Escherichia coli*, which is homologous to human glutathione peroxidase, is not clearly homologous to the other binding proteins. A few extracellular solute-binding receptors fall outside of these eight families and thus may represent new families for which only a single constituent is now known.

Evaluation of the analyses reported has established that high-affinity solute-binding proteins which bind the same or similar solutes but are derived from evolutionarily divergent bacteria are generally more similar in sequence than are proteins specific for very different solutes from the same organism. It is implied from this observation that solute binding imposes severe constraints on the sequence and structure of a protein. Exceptions to the generalization that strikingly similar proteins bind similar substrates include the nickel-binding protein (in the peptide-binding-protein cluster) and iron-binding proteins (in one of the sugar-binding-protein clusters). These surprising findings, as well as the structural and functional divergence of binding proteins during evolution, will be discussed in the concluding sections of this review.

SPECIFICITIES, POLARITIES, AND DOMAIN STRUCTURES OF ABC-TYPE TRANSPORTERS

In all living organisms, solutes are transported across the cytoplasmic membrane of the cell and actively accumulated or extruded in processes driven by ATP hydrolysis (136). One class of transport ATPases consists of multicomponent transport systems with a common structural organization correlating with a unified mechanism of action (8, 35, 37, 38). These transporters, termed ATP binding cassette-type (ABC-type) or traffic ATPases or permeases (9, 75), are functionally diverse. The uptake systems in bacteria transport a wide variety of substrates, including sugars, amino acids, opines, peptides, organic and inorganic phosphate esters, phosphonates, sulfate, polyamines, metallic cations (such as copper, iron, and nickel), molybdenum, organo-iron complexes, and vitamins (see Tables 1 and 6). Bacterial efflux systems of the ABC type exhibit specificity toward drugs such as daunomycin (71), tylosin (165), spiromycin (61, 175), and erythromycin (164); animal toxins such as the *E. coli* hemolysin (50), the *Pasteurella* leukotoxin (79, 193), and the *Bordetella pertussis* cyclolysin (65); the *E. coli* bacteriocin colicin V (64); proteases (111); and capsular polysaccharides and other complex carbohydrates (27, 54, 101, 102, 185, 188). Eukaryotic efflux porters of the ABC type transport sex pheromones in yeasts (103, 125), chloroquine in the protozoan genus *Plasmodium*, methotrexate in *Leishmania* spp. (25, 143), eye pigments in *Drosophila* spp. (44, 140), chloride via the cystic fibrosis transport regulator protein in mammals and in *Xenopus* spp. (161, 198, 205, 225), hydrophobic drugs via the multidrug resistance protein in mammals and *Drosophila* species (28, 48, 68, 69, 221), and peptides across mammalian peroxisomal and endoplasmic reticular membranes (60, 92, 153, 204).

The common protein components of the ABC-type uptake systems include one or two transmembrane proteins which usually span the membrane five or six times each; one or two peripheral membrane ATP-binding proteins localized to the cytoplasmic side of the membrane, which may be anchored to the membrane via a single transmembrane helix; and a high-affinity extracytoplasmic solute-binding protein (75, 86, 96). In gram-negative bacteria, this ligand-specific binding protein is soluble and periplasmic, although its ligand-bound form must associate with the transmembrane constituents of the permease before transport activity can be observed. In gram-positive bacteria, the solute-binding protein is also extracellular but is bound to the membrane via an N-terminal hydrophobic lipid extension (see below). The transmembrane protein components are believed to form solute-specific channels; the peripheral membrane ATP-binding proteins energize the systems and sometimes serve regulatory roles (39), and the ligand-bound binding proteins confer specificity and high affinity for the substrates to the transport systems. Mutant forms of two ABC-type uptake permeases, those specific for maltose and histidine, can function with low efficiency without the periplasmic binding protein (152, 202). The extracellular solute-binding protein constituents of

ABC-type transporters are always lacking if the system catalyzes solute efflux (75).

The domain and subunit organizations of these various ABC-type transporters have recently been reviewed and evaluated (75, 159). The simplest of the bacterial systems are made up of a single integral transmembrane transport protein and a single peripheral membrane ATP-binding protein. Examples of such systems include the DrrAB drug resistance system of *Streptomyces peucetius* (71), the NodIJ system which functions in nodulation in *Rhizobium leguminosarum* (49), and the KpsTM capsular polysaccharide export system of *E. coli* (147). These systems are believed to consist of heterotetramers of the $\alpha_2\beta_2$ type, where the α subunits are the peripheral membrane ATP-binding subunits and the β subunits are the integral membrane channel-forming proteins (159). These export systems lack an extracellular solute-binding protein.

Homologous to these proteins are the BexABC proteins of *Haemophilus influenzae* (102) and the CtrDCB proteins of *Neisseria meningitidis* (54). These three-component capsular polysaccharide efflux transporters possess two nonhomologous integral membrane proteins but only one ATP-binding protein. They may thus have the subunit structure $\alpha_2\beta\beta'$ (159). Other well-characterized ABC-type transporters, such as those responsible for the uptake of histidine in *Salmonella typhimurium* and of maltose in *E. coli*, are known to have the analogous $\alpha_2\beta\beta'\gamma$ structure, where the γ subunit is the periplasmic solute-binding protein (75, 96). Finally, the oligopeptide permease of *S. typhimurium* has an $\alpha\alpha'\beta\beta'\gamma$ structure, with five dissimilar subunits encoded by five distinct genes (75). The corresponding *Bacillus subtilis* oligopeptide transport system similarly consists of an $\alpha\alpha'\beta\beta'\gamma$ pentamer, but surprisingly, and in contrast to the *Salmonella typhimurium* system, the α' subunit (OppF) was found to be dispensable (151, 167). In the absence of the α' (OppF) subunit, the functional permease may assume the structure $\alpha_2\beta\beta'\gamma$.

In addition to the compositional differences noted above, the α and β subunits can be fused to each other in various combinations. For example, in the ribose permease of *E. coli*, the two ATP-binding (α) subunits are fused into a single polypeptide chain roughly twice the size of the normal α subunit (17). The structure of this system is therefore α-$\alpha'\beta\beta'\gamma$, where α-α' indicates a fusion protein consisting of two dissimilar ATP-binding domains. Further, the iron-hydroxamate permease of *E. coli* has the subunit structure of $\alpha_2\beta$-$\beta'\gamma$, where the two integral membrane channel-forming domains are present as a single polypeptide chain (31). In eukaryotes, the α and β subunits can be fused as in the human major histocompatibility complex peptide transporter, which has an $(\alpha\text{-}\beta)_2$ domain/subunit structure (153, 204), or all four domains of the $\alpha_2\beta_2$ heterotetramer can be fused in a single polypeptide chain with or without additional protein domains, as, for example, in the multidrug resistance and cystic fibrosis transport regulator proteins of mammals (see reference 75 for an overview). It is therefore clear that during evolution of ABC-type transporters, gene duplication, gene splicing and/or fusion, and probably operon duplication events have all occurred frequently, leading to the diversity of structural types of these transporters found in nature.

Extracellular solute-binding proteins have never been detected for bacterial efflux systems of the ABC type, and eukaryotic solute-binding proteins homologous to the bacterial proteins have not yet been found. Binding proteins associated with bacterial ABC-type uptake systems are never found fused to the other protein constituents of these systems. Thus, it is clear that although the external solute-binding proteins are essential constituents of the solute uptake systems, they do not play an integral role in the transport process per se. Instead, they probably serve as receptors to trigger or initiate translocation of the solute through the membrane by binding to external sites on the β and β' subunits, thereby inducing conformational changes which are conducive to transport.

EXTERNAL SOLUTE-BINDING LIPOPROTEINS IN GRAM-POSITIVE BACTERIA

The presence of ABC-type transport systems in gram-positive bacteria has only recently been documented. Gilson et al. (63) and Dudler et al. (46) provided evidence for the presence of such systems in *Streptococcus pneumoniae* and *Mycoplasma hyorhinis*. Since these organisms lack an outer membrane and consequently have no periplasm, the solute-specific binding proteins are lipoproteins with an N-terminal glyceride-cysteine, which allows them to be tethered to the external surface of the cell membrane. This hydrophobic anchor maintains the binding protein in the proximity of the external face of the integral cell membrane components of the transport system.

The genes for two such solute-binding lipoproteins were sequenced from *S. pneumoniae* (63). The first, designated MalX, is a maltose-inducible maltodextrin-binding lipoprotein which is 27% identical to the periplasmic MalE-binding protein of *E. coli*. The second, designated AmiA, exhibits 24% identical residues with the oligopeptide-specific periplasmic binding protein OppA of *Salmonella typhimurium* (63, 80). The periplasmic binding proteins of the gram-negative transport systems are essential constituents of maltodextrin and oligopeptide transporters (see below).

The amino-terminal parts of the *E. coli malE* and *S. pneumoniae malX* gene products represent signal peptides characteristic of proteins which are exported through the cytoplasmic membrane. For MalE, it was directly shown that the first 26 residues are cleaved from the remainder of the protein following export (47). In MalX, the region of the potential cleavage site belongs to a well-defined category; it carries a sequence, L-V-A–C-G-S, corresponding to the consensus sequence for the precursors of lipoproteins (L-*Y*-*Z*-cleavage site-C-*y*-*z*, where *Y* is A, S, V, Q or T; *Z* is G or A; *y* is S, G, A, N, Q or D; and *z* is S, A, N, or Q) (reviewed in references 222 and 224). It thus appears likely that the mature MalX protein is a lipoprotein.

Lipoproteins are exported through the cytoplasmic membrane, as are other secreted proteins, but following cleavage the amino-terminal cysteine is thioacylated to give a lipoamino acid. This lipophilic modification is thought to be responsible for the membrane anchorage of a number of exported proteins (135). On the basis of the observed sequence comparisons and the biochemical requirements of the system, membrane attachment of MalX is likely to occur through the same mechanism. MalX would thus be expected to be exposed to the outer face of the membrane as an anchored but otherwise water-soluble protein. It is interesting that a mutant *E. coli* MalE protein which is anchored by its uncleaved amino-terminal signal peptide to the external face of the cytoplasmic membrane can still operate in transport (52).

The situation discussed above is comparable to that of the *amiA* and *oppA* gene products. In *Salmonella typhimurium* OppA, an N-terminal peptide of 23 residues is cleaved from

the remainder of the protein following export (81). For AmiA of *S. pneumoniae*, the region of the potential cleavage site carries the sequence L-A-A–C-S-S, corresponding exactly to the consensus of the lipoprotein precursors.

A gene cluster with the same organization as ABC-type systems from gram-negative bacteria has been described for *M. hyorhinis* (46). The mature extracytoplasmic component, the so-called p37 protein, presents the characteristics of a lipoprotein: the amino-terminal sequence of the mature protein starts with C-S-N-, corresponding to the lipoprotein consensus sequence, and the p37 protein has been shown to be bound to the membrane. The sequence before the potential cleavage site is less typical (A-I-S-cleavage site), but the putative signal sequence is unusual since it contains four phenylalanine residues (46). Since very little is known about signal sequences in mycoplasmas, it is possible that they differ significantly from those in other microorganisms that have been studied.

The considerations mentioned above lead to the proposal that MalX and AmiA in *Streptococcus* species and p37 in *Mycoplasma* species are the functional equivalents of periplasmic solute-binding proteins in gram-negative bacteria. For AmiA, the other components of the transport system are encoded by genes at the *ami* locus (7) and the substrates are likely to be oligopeptides. For MalX, the other components of the transport system have yet to be identified, but the transport substrates are likely to be maltodextrins. For the *Mycoplasma* transporter which includes the p37 protein, two other components, the p29 and p69 membrane proteins, are known but the substrates transported have yet to be identified.

A systematic search for lipoproteins in *Bacillus licheniformis* and *Bacillus cereus* revealed sets of lipoproteins which were released from protoplasts by mild trypsin treatment, suggesting an orientation to the outside of the membrane (135). In addition, the β-lactamases involved in resistance to penicillins are periplasmic proteins in gram-negative bacteria but are present in gram-positive bacteria such as *B. licheniformis* and *Staphylococcus aureus* in substantial amounts in lipoprotein, membrane-bound forms (134). This fact suggests that lipoproteins in gram-positive bacteria play roles equivalent to those of many of the free periplasmic proteins from *E. coli* (135).

Examination of the gene products of the six open reading frames of the *ami* locus of *S. pneumoniae*, *amiABCDEF*, revealed that all but the AmiB protein are homologous to components of the oligopeptide permeases of *Salmonella typhimurium* and *Escherichia coli*. Intriguingly, AmiB was found to be homologous to ArsC, encoding an arsenate reductase which functions in conjunction with the oxyanion pump encoded by the arsenical resistance operon of the R-factor R773 from *E. coli* (29, 73). AmiA is the solute-binding lipoprotein analog of the periplasmic binding proteins of gram-negative bacteria as noted above; AmiC and AmiD are the hydrophobic, transmembrane channel-forming constituents of the system; and AmiE and AmiF are two homologous ATP-binding proteins localized to the cytoplasmic side of the cell membrane, which presumably energize peptide uptake.

Mutations at the *ami* locus of *S. pneumoniae* have pleiotropic effects. Mutants with these mutations were initially isolated on the basis of increased resistance to aminopterin (183), and the mutations were shown to confer sensitivity to an imbalance in the extracellular concentrations of the three branched-chain amino acids, leucine, isoleucine, and valine (182). They also conferred increased resistance to methotrexate (203) and celiptium (2-*N*-methylhydroxyellipticinium) (173). Selective alteration of the transport kinetics for several amino acids was reported for a mutant with a mutation in the *ami* locus and correlated with a decrease in the transmembrane electric potential, deduced from measurement of tetramethylphosphonium accumulation (203). This complex phenotype is not entirely explicable in terms of the assumed function of the *ami* gene products as components of a peptide transport system. Nevertheless, the ability of wild-type strains (but not *ami* mutants) of *S. pneumoniae* to utilize leucine- and arginine-containing peptides to satisfy their needs for these two auxotrophic requirements clearly argues that the Ami transporter recognizes oligopeptides as substrates.

INVOLVEMENT OF EXTERNAL SOLUTE-BINDING PROTEINS IN SENSORY TRANSDUCTION

As noted in the Introduction, the involvement of periplasmic solute-binding proteins as constituents of transport and chemotaxis systems has long been established. Recently, evidence has appeared suggesting that these proteins of gram-negative bacteria as well as homologous lipoproteins of gram-positive bacteria serve as receptors initiating sensory transduction pathways. In this section, this evidence will be summarized.

Phosphate (*pho*) Regulon in *E. coli*

When P_i is available to an enteric bacterium such as *E. coli*, the genes encoded within the phosphate metabolic regulon, responsible for the metabolism of organic phosphate esters and phosphonates, are repressed. Phosphate limitation results in derepression of the more than 30 *pho* genes of the *pho* regulon (210). The principal players in this repression-derepression game are the PhoR and PhoB proteins, which are a sensor kinase-response regulator pair (169, 192) in which PhoR is the sensor and histidyl protein kinase and PhoB is the DNA-binding protein that acts as a transcriptional enhancer. The aspartyl phosphorylated form of PhoB is required for transcription of *pho* regulon genes (117, 118).

Regulation of PhoR activity by P_i requires extracellular P_i and the protein components of the phosphate-specific transport (Pst) system. The latter complex of proteins, including the PstS periplasmic phosphate-binding protein, evidently functions in transmembrane signal transduction. Most (but not all) mutations that abolish transport via the Pst system also abolish transcriptional regulation by PhoB-P, yet repression/derepression is independent of transport per se, since some Pst-specific transport-negative mutants are regulation positive (32, 33) and extracellular P_i, not intracellular P_i, determines the extent of repression. The K_d of PstS for P_i is 1 μM, and 4 μM external P_i is sufficient for full repression, regardless of the internal P_i concentration. It is presumed that the binding of P_i to the PstS protein induces a conformational change in the transmembrane protein complex of the Pst permease (PstABC) and transmits a signal to PhoR, the sensor kinase, possibly through an auxiliary regulatory protein encoded within the *pst* operon, PhoU. PhoU is known to be associated with the inner membrane of the bacterium. The net result would be deactivation of the PhoR kinase and possibly activation of a PhoB-P phosphatase. PhoB would then become dephosphorylated and inactivated, and repression of the *pho* regulon would ensue (210).

Chemical Signaling between *Agrobacterium tumefaciens* and Its Plant Hosts

Agrobacterial species can infect certain plants, giving rise to tumorlike growths. Like the *pho* regulon in *E. coli*, the virulence (*vir*) regulon in *Agrobacterium tumefaciens*, responsible for the pathological state, is transcriptionally regulated by a sensor kinase-response regulator pair, VirA-VirG. These proteins activate transcription of the *vir* regulon in the presence of wound-induced plant phenolic metabolites, such as acetosyringone, as well as sugars which act synergistically with phenolic compounds. The sugar-mediated synergism requires the ChvE protein, a periplasmic protein which binds many sugars (11). Mutants lacking this protein are avirulent on some plant hosts and poorly virulent on others. On binding of its sugar ligand, the ChvE-sugar complex appears to interact with the periplasmic domain of the VirA sensor kinase, which enhances kinase activity. VirA then phosphorylates VirG, and VirG-P activates transcription of the *vir* genes (26, 181).

Many sugars were found to be effective in promoting *vir* gene induction. These sugars included glucose, galactose, and all pyranose sugars with the C-1, C-2, and C-3 hydroxyls in the equatorial position. Acidic monosaccharide derivatives, including D-glucuronate and D-galacturonate, were the strongest inducers among the sugars tested (11). Of the 11 inducing sugars, 8 were identified as known plant metabolites and 7 of the 8 proved to be constituents of major plant cell wall polysaccharides. The results therefore suggested that plant phenolic compounds and monosaccharides released from plant cells and their cell walls function as wound-specific metabolites to signal through the bacterial ChvE-VirA-VirG sensory transduction effector system the presence of the appropriate host environment for invasion (11, 220).

On activation of *vir* gene expression, there is a series of events which results in transfer of a bacterial plasmid, the Ti plasmid, to the nuclei of the infected cells, where it integrates into the plant genome. The plasmid directs overproduction of plant growth hormones which stimulate tumor growth and of opines, which are excreted by the plant to serve as nutrients for the parasitic bacteria. The bacteria must therefore possess uptake systems for the opines. Two such transport systems have been characterized and fully sequenced, one specific for the opine octopine, a reduced condensation product of arginine with pyruvate, and one specific for nopaline, a reduced condensation product of arginine with α-ketoglutarate (100, 200, 226). These transport systems complete a cyclic, three-way signaling system in which (i) the plant first signals its presence by secretion of phenolic compounds and monosaccharides, sensed directly by VirA and the ChvE-binding protein, respectively; (ii) these signals trigger transfer of the Ti plasmid from the bacterium to the plant and integration into the plant genome; and (iii) the Ti plasmid-encoded opine biosynthetic proteins then synthesize opines in the plant cell, which are secreted for uptake via binding protein-dependent transport systems and utilization by the bacterium for energetic and biosynthetic purposes.

Intracellular Animal Pathogenesis Involving *Salmonella typhimurium*

Enteric bacteria possess three known genetically distinct peptide uptake systems with overlapping substrate specificities. The first system, the oligopeptide permease (Opp), mediates the uptake of peptides containing up to five amino acid residues with practically no specificity for the nature of the amino acid chains (72, 80, 148). This system plays an essential role in the recycling of peptides released from the cell wall during growth (66). The second system, a dipeptide permease (Dpp), like Opp, is a periplasmic-binding-protein-dependent transport system, but its specificity is limited to dipeptides. The third system, a tripeptide permease (Tpp), is less well characterized but has greatest affinity for hydrophobic tripeptides. Expression of the *tpp* genes is induced during anaerobic growth, but the physiological role of the Tpp system is still unclear (62, 91).

The genes encoding a fourth presumed peptide transport system from *Salmonella typhimurium* of the ABC type have recently been sequenced (67a). The proteins encoded by these genes show striking sequence similarity to the components of the known peptide transporters of enteric and gram-positive bacteria. A genetic defect in this transport system results in loss of virulence in mice. These observations suggest that, as in *Bacillus* species (see below), peptide transport may be required for sensing alternative physiological conditions; for this gram-negative bacterium, this ability is required for maintenance of the pathological state. Alternatively, the proteins of this novel ABC-type transport system may function directly in signal transduction, as appears to be the case for the Pst system of *E. coli* and the ChvE-binding protein of agrobacteria.

Control of Sporulation versus Alternative Starvation-Induced States in *Bacillus subtilis*

Recently, the genes encoding the protein components of an oligopeptide transport system in *B. subtilis* were sequenced and the encoded system was shown to play a role in the initiation of sporulation (151, 167). A mutation, originally designated *spo0K*, giving rise to a stage zero phenotype (82, 83) was found to map within the five-cistron operon which encodes the transport system. All five encoded proteins were found to be highly homologous to the corresponding components of the oligopeptide transporters of *Salmonella typhimurium* and *E. coli*.

Studies with toxic peptide analogs provided evidence that this operon does indeed encode a peptide transport system. The deduced amino acid sequences of the five open reading frames were very similar to the OppA, OppB, OppC, OppD, and OppF proteins of *Salmonella typhimurium*. The OppA solute-binding protein was the least conserved, especially at the amino terminus, with 182 identical (33%) and 230 conserved (42%) residues compared with the equivalent *Salmonella typhimurium* protein. The ATP-binding proteins OppD and OppF were more highly conserved, with 165 identical (54%) residues. The hydrophobic OppB and OppC proteins fell between these two extremes. The degree of conservation in all cases was high and allowed unequivocal identification of these sequences as the genes of the *B. subtilis opp* operon. Interestingly, unlike the Opp system of *Salmonella typhimurium*, one of the two *B. subtilis* ATP-binding proteins, OppF, was not required for peptide transport or sporulation, although all the other protein constituents of the transport system were required. The fact that OppF was dispensable led to the possibility that OppD can form a functional homodimer, which eliminates the need for OppF altogether. Surprisingly, OppF has been reported to be required for genetic competence (167).

As expected, the OppA peptide-binding protein had a signal sequence characteristic of lipoproteins with an amino-

terminal lipoamino acid anchor. Cellular location studies revealed that OppA of *B. subtilis* was associated with the cell during exponential growth but was released into the medium in the stationary phase. The OppA protein was shown to be required for transport of peptides via this system. Whether release into the medium serves a biological function or whether the protein can be recaptured by the same or other cells following its release has not been clarified. It is possible that modification to the lipoprotein derivative provides a mechanism for regulating retention and release of the protein from the cell, a process which might be directly or indirectly important to the regulation of sporulation.

In this connection, it has been postulated that the accumulation of peptides derived from the degraded cell wall (the peptidoglycan) plays a signaling role in the initiation of sporulation and that the sporulation defect in *opp* mutants results from an inability to transport these peptides (151). The rationale for this notion is as follows. Peptide-like molecules probably play a general role in signaling differentiation in all sporulating organisms (70). It is possible that small peptides composed of protein amino acyl residues linked exclusively by α-peptide bonds do not serve a signaling role because they are be rapidly degraded by intracellular peptidases. However, Opp in *Salmonella typhimurium* and *E. coli* can transport peptidase-resistant cell wall peptides released from the peptidoglycan during growth (66), and the same would be expected of the gram-positive homolog. It is therefore possible that cell wall peptides serve to signal the onset of the stationary phase and initiate sporulation. Since *opp* mutants cannot recycle cell wall peptides, the peptides do not accumulate internally and this presumed sporulation signal is not generated. An association of murein components with sporulation in *Myxococcus xanthus* has also been reported (180).

Recently, a dipeptide transport system, expressed early during sporulation in *B. subtilis* and dependent on the *spo0A* gene product for expression, has been characterized (121, 184). The genes encoding the five protein constituents of the *dciA* operon (*dciAA* through *dciAE*) were found to include homologs of the genes encoding the Opp systems of both *B. subtilis* and enteric bacteria as well as those encoding the Dpp system of *E. coli* (141; see below). Interestingly, the DciA system of *B. subtilis*, which clearly catalyzes dipeptide uptake, does not interfere with sporulation when defective. Its function has been proposed to involve facilitation of the adaptation from nutrient-rich to nutrient-poor conditions (121). Thus, it may actually antagonize sporulation by providing an alternative route to sporulation; i.e., it and the products of many other starvation-induced genes may function to improve the nutritional state of the cell both by generating new nutrients (through the generation of antibiotics and extracellular macromolecular degradative enzymes) and by improving the ability of the cell to utilize meager supplies of limiting nutrients (through the induction of high-affinity transport systems).

The five proteins encoded within the *dciA* operon, DciAA to DciAE, were screened against the data base for sequence similarity with other sequenced proteins. The protein encoded by the first open reading frame, DciAA, did not exhibit significant sequence similarity to any other sequenced protein. DciAB was a hydrophobic protein (six putative transmembrane segments) which exhibited 53% identity with OppB of *B. subtilis* and 42% identity with OppB of *Salmonella typhimurium*. DciAC was similar to DciAB in being hydrophobic, and it most closely resembled the OppC proteins of *B. subtilis* and *Salmonella typhimurium*. DciAD proved to be an ATP-binding protein homologous to OppD and OppF. Finally, DciAE resembled the external dipeptide- and oligopeptide-binding proteins. This last protein was shown to be essential for dipeptide utilization, although tripeptides could still be utilized by strains lacking the protein. The tripeptides were apparently utilized via the *Bacillus* Opp system (121). Evidence for a tripeptide-specific transporter in *B. subtilis*, analogous to the corresponding enteric system, has not yet been reported.

COMPUTER METHODS FOR DEFINING EVOLUTIONARY AND STRUCTURAL RELATIONSHIPS OF EXTERNAL SOLUTE-BINDING PROTEINS

The Protein Information Resource (PIR) of the National Biomedical Research Foundation (NBRF) data base (version 32) and the current GenBank (version 71.0), EMBL (version 24.0), and SwissProt (version 21.0) data bases were screened in all homology searches by using the FASTA program (112, 149). Significance was evaluated by calculating the binary comparison scores, expressed in standard deviations (SD), by using the RDF2 program with Ktuple=2 (149). A comparison score of 8 to 10 SD was considered to reflect a sufficient degree of sequence similarity to establish homology (36, 41, 42). Both the ALIGN (36) and Los Alamos (93) programs were used to confirm significance. These programs introduce comparable gaps to maximize identities when comparing protein sequences as well as when comparing randomized sequences. Comparison scores are recorded in SD higher than obtained with 200 comparisons of randomized sequences of the two protein segments analyzed.

Homologous protein sequences were aligned by using the Needleman-Wunsch algorithm, and relative evolutionary distances and phylogenetic trees were determined by using the progressive alignment or gap-minimizing TREE program of Feng and Doolittle (51) and Doolittle and Feng (43). All sequence analyses and data base searches were performed by using the GCG package from the University of Wisconsin (40) and the University of California at San Diego VAX/VMS DNASYSTEM package (186). The figures presenting multiple sequence alignments show selected regions of homologous proteins which generally exhibit extensive sequence similarity to document the relatedness of these proteins; the complete multiple alignments were generated for all clusters of proteins and are available from M. H. Saier, Jr., on request.

Signature sequences (14, 15) for the eight clusters or families of binding proteins were determined from the most highly conserved regions of the multiply aligned proteins. These alignments are shown in the figures; for the derived signature sequences, see Table 4. Each of these signature sequences was screened against the protein data banks (PIR version 32 and SwissProt version 22). In no case did a protein other than the members of the cluster for which that sequence was derived exhibit this identifying sequence. Thus, each of the signature sequences listed in Table 4 is family specific and can assist in the identification of new members of each of these families.

EIGHT FAMILIES OR CLUSTERS OF BACTERIAL SOLUTE-BINDING PROTEINS

Table 1 lists, characterizes, and provides standard abbreviations for 45 of the 52 bacterial solute-binding proteins included in this study as well as for a homologous transcriptional regulatory protein. The remaining seven binding pro-

TABLE 1. Eight families (clusters 1 to 8) of periplasmic binding proteins of gram-negative bacteria, homologous lipoproteins of gram-positive bacteria, and homologous DNA-binding transcriptional regulators[a]

Abbreviation[b]	Specificity	Gene	Interactive transport system	No. of residues[c]	Organism	Accession no.	Reference
Cluster 1							
MalEco	Maltose	*malE*	MalFGK	396	*E. coli*	PIR A03428	47
MalSty	Maltose	*malE*	MalFGK	396	*S. typhimurium*	PIR SO5331	34
MalKae	Maltose	*malE*	MalFGK	396	*K. aerogenes*	PIR S05328	34
MalSpn	Maltose	*malX*	MalMP	402	*S. pneumoniae*	GenBank J01796	105
GlpEco	Glycerol 3-phosphate	*ugpB*	UgpACE	438	*E. coli*	PIR S03780	145
MsmSmu	Multiple sugars	*msmE*	MsmKFG	420	*S. mutans*	GenBank M77351	168
SfuSma	Iron	*sfuA*	SfuBC	338	*S. marcescens*	GenBank M33815	10
FbpNme	Iron	*fbp*	—[d]	330	*N. meningitidis*	GenBank X53467	19
				390 ± 37			
Cluster 2[e]							
RibEco	Ribose	*rbsB*	RbsACD	296	*E. coli*	PIR A03425	17
RibSty	Ribose	*rbsB*	RbsACD	296	*S. typhimurium*	PIR A03426	23
AraEco	Arabinose	*araF*	AraGH	306	*E. coli*	PIR A03424	84
GalEco	Galactose	*mglB*	MglACE	309	*E. coli*	PIR A03427	115
GalCfr	Galactose	*mglB*	—	332	*C. freundii*	GenBank P23925	Unpublished
ChvAtu	Multiple sugars	*chvE*	None[f]	246	*A. tumefaciens*	GenBank M30318	88
				291 ± 26			
Cluster 3							
HisSty	Histidine	*hisJ*	HisQMP	260	*S. typhimurium*	PIR A03407	77
LAOSty	Lysine-arginine-ornithine	*argT*	HisQMP	260	*S. typhimurium*	PIR A03408	76
GlnEco	Glutamine	*glnH*	GlnPQ	248	*E. coli*	PIR S03181	138
GlnBst	Glutamine	*glnH*	GlnPQ	262	*B. stearothermophilus*	GenBank M61017	223
OccAtu	Octopine	*occT*	OccQMP	276	*A. tumefaciens*	GenBank M77784	206
NocAtu	Nopaline	*nocT*	NocQMP	283	*A. tumefaciens*	GenBank M77785	226
BaaNgo	Basic amino acids	*p29*[g]	—	268	*N. gonorrhoeae*	GenBank X64421	108
				265 ± 12			
Cluster 4							
LeuEco	Leucine	*livK*	LivHMGF	369	*E. coli*	PIR B23576	106
LIVEco	Leucine-isoleucine-valine	*livJ*	LivHMGF	344	*E. coli*	PIR A03415	4
LIVCfr	Leucine-isoleucine-valine	*livJ*	—	367	*C. freundii*	PIR S14619	Unpublished
LIVPae	Leucine-isoleucine-valine	*braC*	BraDEFG	373	*P. aeruginosa*	PIR A36125	85, 86
AmiPae[h]	"[AmiC, transcriptional repressor]"	*amiC*	None[f]	385	*P. aeruginosa*	PIR A40359	219
				368 ± 15			
Cluster 5							
OppEco	Oligopeptide	*oppA*	—	543	*E. coli*	GenBank J05433	94
OppSty	Oligopeptide	*oppA*	OppBCDF	542	*S. typhimurium*	GenBank X04194	80
OppSpn	Oligopeptide	*amiA*	AmiCDEF	493	*S. pneumoniae*	GenBank X17337	7
DppBsu	Dipeptide	*dciAE*	DciAABCD	543	*B. subtilis*	GenBank X56678	121
OppBsu	Oligopeptide	*oppA*	OppBCDF	545	*B. subtilis*	GenBank X56347	151, 167
DppEco	Dipeptide	*dppA*	—	535	*E. coli*	GenBank M35045	1
NikEco	Nickel	*nikA*	—	524	*E. coli*	Not available	L.-F. Wu, personal communication
XP55Sli	Not known	*xp55*[g]	—	542	*S. lividans*	GenBank Y00142	24
				533 ± 18			
Cluster 6							
SulEco	Sulfate	*sbp*	CysTWA	329	*E. coli*	PIR B25206	74
SulSty	Sulfate	—[g]	—	310	*S. typhimurium*	PIR A03403	89
SulSsp	Sulfate	*sbpA*	CysTWR	350	*Synechococcus* sp.	GenBank M65247	107
TsuEco	Thiosulfate	*cysP*	CysTWA	338	*E. coli*	GenBank M32101	87
[PhoEco][i]	Phosphate	*phoS*	PstCAB	346	*E. coli*	PIR A30277	194
				335 ± 16			
Cluster 7							
DcaRca	Dicarboxylate	*dctP*	—	333	*R. capsulatus*	Not available	178
[CitSty][i]	Citrate	*tctC*	—	303	*S. typhimurium*	Not available	215
							W.W. Kay, personal communication

Continued on following page

TABLE 1—*Continued*

Abbreviation[b]	Specificity	Gene	Interactive transport system	No. of residues[c]	Organism	Accession no.	Reference
				318 ± 21			
Cluster 8							
FecEco	Iron-enterobactin	*fecB*	FecCDE	300	*E. coli*	GenBank M26397U01	189
FepEco	Iron-dicitrate	*fepB*	FepCDG	318	*E. coli*	GenBank M29730	179
FhuEco	Iron-hydroxamate	*fhuD*	FhuBC	296	*E. coli*	GenBank X05810	31, 99
FatVan	Iron-anguibactin	*fatB*	FatCD	322	*V. anguillarum*	GenBank M74068	2
				309 ± 13			
[BtuEco][i]	Vitamin B_{12}	*btuE*	BtuCD	183	*E. coli*	GenBank M14031	53

[a] The abbreviations used in this study for the various proteins are provided. These usually indicate the specificity of the binding protein (first three letters) and the organism from which it was obtained. The genus and species are indicated by the second three letters of the abbreviation. In some cases, such as the iron- and iron-complex-binding proteins as well as the ChvE sugar-binding protein of *A. tumefaciens*, the gene designations are used instead of the three-letter designation of the solute bound. DNA-binding proteins homologous to cluster 2 binding proteins are not listed. Sequenced periplasmic binding proteins which do not fall within the eight clusters of proteins defined in this study are not included in the table. They are listed in the text.

[b] Proteins are listed according to the cluster of homologous proteins in which they are found as reported in this study.

[c] Number of residues in the (precursor) protein encoded by the structural gene for that protein. The mean number of residues for all of the solute-binding proteins within an indicated cluster (± SD) is provided below the list of those proteins. Values for individual subclusters are given in Table 3 when proteins of nonuniform molecular size are included within a single cluster.

[d] —, the sequences of these genes are not yet available.

[e] Cluster 2 periplasmic binding proteins are homologous to a large group of DNA-binding proteins as described previously (207).

[f] This binding protein does not interact with a transport system.

[g] This binding protein does not have a gene designation.

[h] AmiC is not a periplasmic binding protein but a cytoplasmic transcriptional regulatory protein (219).

[i] Proteins in brackets are not demonstrably homologous to other members of the cluster to which these proteins have been assigned, because the comparison scores recorded in Table 2 are too small. However, common function, similar binding specificities, similar size, and elevated similarity scores obtained when compared with other members of their cluster all argue in favor of homology (42, 119).

teins are described with references to their sequences in the section on additional binding-protein families, below. All of these proteins were screened against the current data bases to establish that all currently available homologous proteins were identified. No eukaryotic protein was found in these searches, with the sole exception of the vitamin B_{12}-binding protein of *E. coli*, which is homologous to human glutathione peroxidase (see the section on cluster 8 binding proteins, below). Extracytoplasmic binding proteins that did not exhibit significant sequence similarity to any of the proteins in Table 1 or to any other protein in the data base were largely excluded from our study. These proteins included the glycine betaine-binding proteins, ProX, of *E. coli* and *Salmonella typhimurium* (67, 144, 190), the PhnD phosphonate-binding protein of *E. coli* (30, 116, 127, 211), the PotD polyamine-binding protein of *E. coli* (58), and the p37 protein of *M. hyorhinis*, which is believed to be a solute-binding protein (46). Another periplasmic binding protein which was excluded from our study was the small (91-residue) mercuric-ion-binding protein, MerP, of gram-negative bacteria (139), which was not homologous to any of the proteins listed in Table 1. It was found to be homologous to what appear to be cation-binding domains of bacterial mercuric reductases, of the cadmium ATPase of *Staphylococcus aureus*, and of the cation ATPase of *Rhizobium meliloti* (49a). Other excluded proteins are mentioned below.

The 46 proteins listed in Table 1 were grouped into eight clusters based on their sequence similarities, and these groupings were generally found to correlate with the molecular sizes and solute-binding specificities of the included proteins. Selected statistical analyses (binary comparisons) are reported in Table 2. Every protein within a particular cluster is included in the analysis reported in Table 2 in order to document homology. On the basis of the superfamily concept (42), the proteins within a cluster are shown to be homologous (comparison scores of ≥8 to 10 SD [41, 42]) with the following exceptions: cluster 2, AraEco; cluster 6, PhoEco; cluster 7, CitSty; cluster 8, BtuEco. Despite the lower comparison scores reported for these four proteins, the values obtained for each of these proteins with other members of their assigned cluster (6 to 8 SD) suggest that they may make up part of the cluster to which they have been assigned. This assignment is sometimes substantiated by the common functions for, similar binding specificities of, and similar sizes of the compared proteins (42, 119). Limited regions of striking sequence similarity were sometimes detected when making intercluster comparisons, but in no case was the degree of similarity as extensive as the intracluster similarities noted in Table 2.

The number of amino acid residues in each of these proteins is listed in Table 1. It can be seen that, with only a few exceptions, there is little size variation within a particular cluster. The only exceptions are the two iron-binding proteins in cluster 1, the ChvE protein (ChvAtu) in cluster 2, and the BtuE protein (BtuEco) in cluster 8 (Table 3). Excluding these proteins, the size ranges (numbers of residues) in each of the clusters are very restricted, as summarized in Table 3. The eight clusters are presented and discussed below.

Cluster 1 Binding Proteins Specific for Malto-Oligosaccharides, Multiple Sugars, α-Glycerol Phosphate, and Iron

Cluster 1 proteins, specific for oligosaccharides, α-glycerol phosphate, and iron, make up two size clusters (Tables 1 and 3), the larger carbohydrate-binding proteins (396 to 438 residues) and the smaller iron-binding proteins (330 to 338 residues). Figure 1A shows a similarity plot for the entire multiple alignment, and Fig. 1B shows the segment of the

TABLE 2. Binary comparison scores for selected intracluster pairs of proteins[a]

Cluster and aligned pair	No. of residues compared[b]	% ID[c]	Comp. score[d]	Probability[e]
Cluster 1				
MalEco & MalKae	396	94	84	$<10^{-35}$
MalEco & MalSty	396	93	75	$<10^{-35}$
MalEco & MalSpn	369	32	17	$<10^{-35}$
MalSpn & MsmSmu	357	28	9	10^{-19}
MsmSmu & GlpEco	330	17	10	10^{-23}
MalEco & SfuSma	40	32	9	10^{-19}
SfuSma & FbpNme	317	38	28	$<10^{-35}$
Cluster 2				
RibEco & RibSty	64	97	26	$<10^{-35}$
RibEco & ChvAtu	64	23	10	10^{-23}
RibEco & GalEco	64	30	10	10^{-23}
RibEco & AraEco	64	22	7	10^{-12}
GalEco & GalCfr	309	96	127	$<10^{-35}$
Cluster 3				
LAOSty & HisSty	84	73	26	$<10^{-35}$
LAOSty & GlnEco	83	33	12	10^{-25}
LAOSty & OccAtu	84	42	16	$<10^{-35}$
LAOSty & NocAtu	84	45	19	$<10^{-35}$
LAOSty & BaaNgo	84	37	12	10^{-25}
Cluster 4				
LeuEco & LIVEco	128	85	37	$<10^{-35}$
LeuEco & LIVPae	128	66	34	$<10^{-35}$
LIVPae & AmiPae	339	19	17	$<10^{-35}$
LIVEco & LIVCfr	343	94	166	$<10^{-35}$
Cluster 5				
OppSty & OppEco	543	84	160	$<10^{-35}$
OppEco & DppBsu	476	32	49	$<10^{-35}$
OppEco & OppBsu	547	32	66	$<10^{-35}$
OppEco & DppEco	515	22	26	$<10^{-35}$
OppEco & OppSp	290	24	20	$<10^{-35}$
OppEco & NikEco	498	23	23	$<10^{-35}$
XP55Sli & OppSty	312	25	24	$<10^{-35}$
Cluster 6				
SulEco & SulSty	311	91	80	$<10^{-35}$
SulEco & SulSsp	134	52	36	$<10^{-35}$
SulEco & TsuEco	227	53	53	$<10^{-35}$
SulEco & PhoEco	55	16	6	10^{-9}
Cluster 8				
FecEco & FepEco	208	29	11	10^{-25}
FecEco & FhuEco	253	19	8	10^{-15}
FecEco & FatVan	284	20	12	10^{-25}
FatVan & FhuEco	264	16	13	10^{-25}
FecEco & BtuEco	80	24	6	10^{-9}

[a] All proteins included in the phylogenetic trees presented in this paper are listed with comparison scores which in most instances establish (but in one instance [AraEco] only strongly suggests) homology. Values of 8 to 10 SD for protein domains are considered sufficient to establish homology (41, 42).

[b] Number of residues in the overlapping segments compared within the two proteins.

[c] %ID, percent identity within the two aligned sequences being compared.

[d] The comparison or alignment scores (comp. score) were determined with the RDF2 program (Ktup2) with 200 random shuffles (112, 149).

[e] The probability that the observed degree of similarity arose during evolutionary history by chance is based on the comparison score (36).

multiple alignment corresponding to the bar in Fig. 1A which is relatively well conserved. Although this region of the multiple alignment shows the highest degree of similarity for these sequences, the divergence is still quite considerable. The short segment of the multiple alignment of the eight proteins which make up cluster 1 shown in Fig. 1B was selected to derive a signature sequence (14, 15) for this family of proteins (Table 4). It can be seen from the results summarized in Table 5 that the cluster 1 proteins exhibit a similarity score of 0.45, well above the control value of 0.23.

The proteins in cluster 1 are diverse, exhibiting specificity for malto-oligosaccharides (Mal), multiple oligosaccharides (MSM), iron (Sfu and Fbp), and α-glycerol phosphate (Glp). In some regions excellent alignment between some of the proteins but not others was observed, whereas in other regions good alignment between other proteins was observed. For example, MalEco aligns particularly well with SfuSma in the central portion of the region shown in Fig. 1B. The complete multiple alignment revealed numerous short regions exhibiting degrees of sequence similarity approaching that shown in Fig. 1.

The cluster 1 phylogenetic tree is shown in Fig. 1C. It can be seen that the malto-oligosaccharide-binding proteins of enteric bacteria (*E. coli*, *Salmonella typhimurium*, and *Klebsiella aerogenes*) are very closely related (93 to 95% identity throughout their lengths). The partially sequenced maltodextrin-binding protein of *S. pneumoniae* (MalSpn) is more distant but nevertheless exhibits 32% identity with the *E. coli* MalE (MalEco)-binding protein throughout most of its length (Table 2). The multiple-oligosaccharide-binding protein of *Streptococcus mutans* (MsmSmu), the α-glycerol-phosphate-binding protein of *E. coli* (GlpEco), and the iron-binding proteins of *Serratia marcescens* and *N. meningitidis* (SfuSma and FbpNme, respectively) form three distant branches of the tree (Fig. 1C).

The periplasmic iron-binding proteins of cluster 1 are the SfuA iron-binding protein of *Serratia marcescens* (10) and iron-binding proteins of *Neisseria gonorrhoeae* and *N. meningitidis* (19, 128). The two *Neisseria* proteins exhibit nearly 100% identity with each other, and consequently only one is included in this study. These two proteins exhibit 38% identity in an overlapping region of 317 amino acid residues with the SfuA protein. The SfuA protein gave a comparison score of 9.3 SD ($P = 10^{-20}$) with the MalE-binding protein of *E. coli* (18% identity over an overlapping region of 224 residues, or 32% identity over an overlapping region of 40 residues). This comparison score establishes a common evolutionary origin for these proteins and therefore places the iron-binding proteins in cluster 1.

Cluster 2 Binding Proteins Specific for Hexoses and Pentoses

Cluster 2 proteins, specific for hexoses and pentoses, are fairly uniform in size (296 to 306 residues), with the ChvE protein representing the only exception (246 residues [Tables 1 and 3]). Because the galactose-binding proteins of *E. coli* and *Citrobacter freundii* are nearly identical (Table 2), only the former protein was included in this study. As revealed by the similarity plot shown in Fig. 2A, cluster 2 proteins are substantially less divergent from each other than are cluster 1 proteins (Table 5). The N-terminal parts of these proteins are more strongly conserved than the C-terminal domains are. The portion of the complete multiple alignment for the five cluster 2 sugar-binding proteins shown in Fig. 2B, corresponding to the region indicated by the bar in Fig. 2A, also reveals a greater degree of sequence similarity than

TABLE 3. Size comparisons among the binding proteins[a]

Cluster	No. of residues[b]	Mean no. of residues[c]	Specificity
	91	91	Mercuric ion (MerP)
8B	183	183	Vitamin B_{12} (BtuE)
2B	246	246	Multiple sugars (ChvE)
3	248–283	265 ± 12	Polar amino acids
2A	296–332	308 ± 15	Pentoses and hexoses
8A	296–322	309 ± 13	Iron complexes
	309	309	Glycine betaine (ProX)
7	303–333	318 ± 21	Organic anions
1B	330–338	334 ± 6	Iron (Sfu and Fbp)
6	310–350	335 ± 16	Inorganic anions
	338	338	Phosphonates, phosphite, phosphate, and phosphate esters (PhnD)
	348	348	Polyamines (PotD)
4	344–385	368 ± 15	Aliphatic amino acids and amides
1A	396–438	408 ± 17	Oligosaccharides and α-glycerol phosphate
	436	436	Copper (putative) (NosD)
5	493–543	533 ± 18	Peptides and nickel

[a] Proteins not included within a cluster are discussed in the text, and references are provided therein.
[b] Number of residues in the native (uncleaved) protein, or range of numbers of residues in the proteins when more than one is present in the designated cluster (Table 1).
[c] Variation within each cluster of proteins is provided as standard deviation.

observed for the cluster 1 binding proteins. Within this gapless segment of 38 overlapping residues, residues in five positions are fully conserved. The signature sequence for this cluster of proteins, derived from the multiple alignment shown in Fig. 2B, is given in Table 4. Even within the depicted region, which represents the most highly conserved portions of these sequences, divergence is very considerable. Despite the greater divergence of the arabinose-binding protein of *E. coli* (AraEco) relative to the other members of cluster 2 (Table 2), this protein clearly exhibits a striking degree of sequence similarity with the other four proteins and is unquestionably a member of this cluster (i.e., 32% identity with ChvAtu for the segment shown [209]).

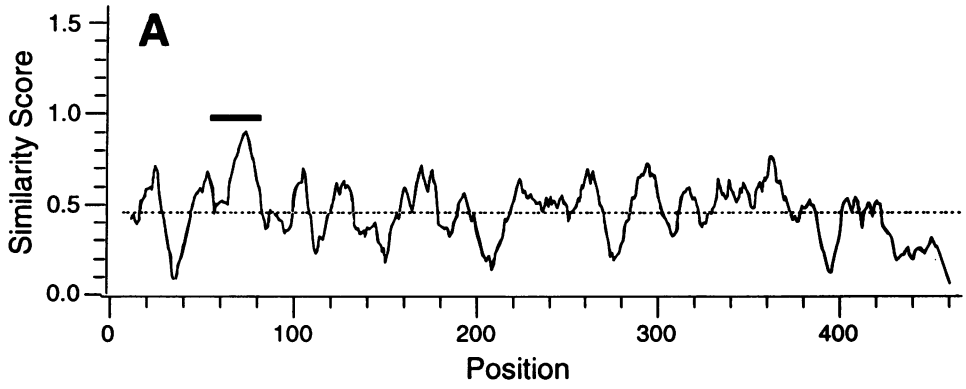
The cluster 2 tree (Fig. 2C) reveals four branches, with the ribose-binding proteins of *E. coli* and *Salmonella typhimurium* nearest the center of the unrooted tree. This fact is particularly noteworthy since the ribose-binding proteins exhibit the greatest degree of sequence similarity to a family of transcriptional regulatory proteins including the lactose, galactose, and fructose repressors of *E. coli* (207). It is interesting that the ChvE protein of *A. tumefaciens* (ChvAtu) serves as a receptor for communication between the bacterial parasite and the host plant in a complex relationship which promotes plant cell tumorigenesis (206).

Limited regions of sequence similarity were noted when some of the cluster 1 binding proteins were compared with some of the cluster 2 binding proteins. Comparison scores of 7 SD (GlpEco versus GalEco) and 6 SD (MalEco versus GalEco) were obtained, for example, for short regions between members of these two clusters. The implications of these observations are not clear, since these comparison scores are insufficient in magnitude to establish homology and the segments compared were short.

Cluster 3 Binding Proteins Specific for Polar Amino Acids and Opines

Cluster 3 proteins, specific for polar amino acids and opines, exist in the size range of 248 to 283 residues (Tables 1 and 3). Like those for cluster 1 and cluster 2 proteins, the similarity plot for cluster 3 proteins (Fig. 3A) reveals regions of strong conservation in the N-terminal domains. The average similarity score is substantially greater than those for cluster 1 and cluster 2 proteins (Table 5). A selected portion of the multiple alignment for the six cluster 3 binding proteins, corresponding to the bar in Fig. 3A, is shown in Fig. 3B. Three of the proteins shown bind basic amino acids, one binds glutamine, and two bind opines. Opines arise by the reductive condensation of a basic amino acid such as arginine with a keto acid such as pyruvate, and consequently they are structurally related to the polar amino acid ligands of the other cluster 3 binding proteins. Within the alignment of 57 residues, shown in Fig. 3B, residues in nine positions are fully conserved. In agreement with the large comparison scores recorded in Table 2, these proteins exhibit very substantial sequence similarity. Two signature sequences, derived from the segment shown in Fig. 3B, are presented in Table 4.

The degree of sequence similarity observed for this cluster of proteins is reflected in the shorter branch lengths of the phylogenetic tree (Fig. 3C). Three major branches are seen with the two basic amino acid-binding proteins of *Salmonella typhimurium* and the glutamine-binding protein of *E. coli* making up one branch, the basic amino acid-binding protein of *N. gonorrhoeae* forming the second branch, and the opine-binding proteins of *A. tumefaciens* representing the third branch. The histidine- and the lysine-arginine-ornithine (LAO)-binding proteins of *Salmonella typhimurium*, which cluster tightly together, are known to interact with the same transport system (76). The recently sequenced glutamine-binding protein of *Bacillus stearothermophilus* (Table 1) is strikingly similar to the basic amino-acid-binding protein of *N. gonorrhoeae* and the glutamine-binding protein of *E. coli* (Table 2). A fragment of the LAO-binding protein of *E. coli* has been sequenced but was not included in this study since the region sequenced is nearly identical to the corresponding region of the *Salmonella typhimurium* protein.



B

```
                  4 3 4 5 3 5 3 7 5 4 4   5 6 5 8 6 4 4 3 3 3 4 3 4 5
Mal Eco   (46) L A E V G K K F E K D   T G I K V T V E H P D K L E (70)
Mal Kae   (46) L A E V G K K F E K D   T G I K V S V E H P D K L E (70)
Mal Sty   (46) L A E V G K K F E Q D   T G I K V T V E H P D K L E (70)
Mal Spn   (57) I E E V A K A Y E K E   A G V K V T L K T G D A L G (81)
Msm Smu   (51) L K K I I K D F E R E   N P   K I H V K M T S V P D (74)
Sfu Sma   (44) V K S W V D G F T K D   T G I K V T L R N G G D S E (68)
Fbp Nme   (35) A Q A V A D A F T R A   T G I K V K L N S A K G D Q (59)
Glp Eco   (41) V D S L A Q R F N A E N P D Y K I V P T Y K G N Y E (66)
```

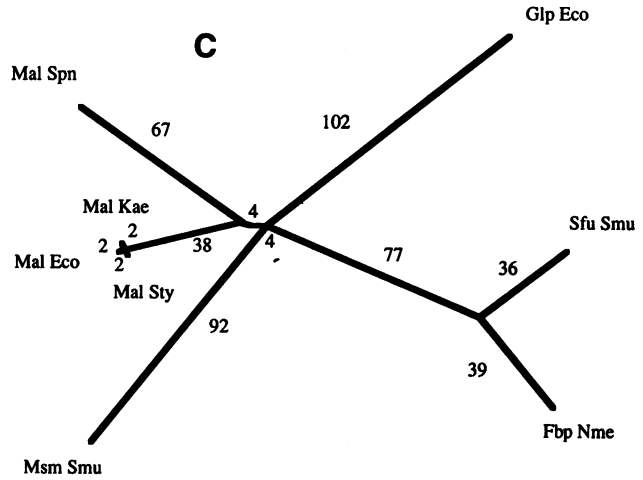


FIG. 1. Analyses of cluster 1 (oligosaccharide-, α-glycerol phosphate-, and iron-binding) proteins. Abbreviations are as indicated in Table 1. (A) The similarity score at a particular position was calculated for a sliding window of 10 residues by averaging the similarity scores obtained for those 10 residue positions. The average similarity across the entire alignment is indicated by the dashed line. The bar above the graph represents the part of the multiple alignment which is shown in panel B. (B) One segment of the cluster 1 protein multiple alignment. The number provided above each aligned residue indicates the number of residues in which the most frequently occurring residue is found at this position. Residues shown in boldface type are fully conserved. The residue numbers within each protein segment are provided at the beginning and end of each sequence. The complete multiple alignment of the cluster 1 binding proteins as well as those of the other clusters of proteins (clusters 2 to 8) are available from M. H. Saier, Jr., on request. (C) Phylogenetic tree for the proteins that make up cluster 1. Relative branch distances (drawn to scale) are provided adjacent to the branches. Comparison scores establishing intracluster homology are presented in Table 2.

Cluster 4 Binding Proteins Specific for Aliphatic Amino Acids and Amides

Cluster 4 proteins bind aliphatic amino acids and amides and have a size range of 344 to 385 residues (Tables 1 and 3). The leucine-isoleucine-valine (LIV)-binding proteins of *E. coli* and *C. freundii* are nearly identical, and consequently only the former was selected for study. The similarity plot shown in Fig. 4A for these binding proteins reveals greater sequence similarity in the N-terminal domains than in the C-terminal domains, and the average similarity score for the entire alignment is larger than those obtained for clusters 1 to 3 (Table 5). A large segment (88 residues) of the multiple alignment for the four cluster 4 proteins is shown in Fig. 4B, and a signature sequence for these proteins is provided in Table 4.

TABLE 4. Signature sequences for the eight binding-protein families[a]

Family	Signature sequence[b]	No. of positions specified
1	(LIVA)X_2(LIVW)(GAIV)(KDQ)X(FY)(ETN)X(DEA)$X_{(0,\ 1,\ or\ 2)}$(GND)(IVPY)K(VI)X(VLP)	12
2	K(LIVFAG)$_3$IX_3D(SGP)X_3(GS)X(LIVA)$_2$$X_2$A	10
3A[c]	GF(DE)(LIV)DLX_3(LIVM)(CA)(KE)	10
3B[c]	P(SA)LX_2(KG)X_2D(LIVMA)$_3$(SA)(GAS)(LIVM)	11
4	GAX_2A(LIV)X_2(LIV)NX_2GG(LIV)XG	10
5	(LIVM)AX_2(WI)$X_{1\ or\ 2}$(SN)(KE)DX_4T(FY)X(LIV)RX_3K	10
6A[c]	GNPKXIXDWXDLX(RK)X_2VX(LIV)$_2$	13
6B[c]	NPKXSGXARX_2(FY)L(AG)AWG	12
7	SX(AG)KFGX_2GVXD	8
8	(EQ)P(SN)(LMA)EX(LIV)X(AE)X(MK)P(DS)(LF)(IVM)(LIV)	13

[a] Each signature sequence was screened and shown to be specific for the binding-protein cluster indicated as described in Computer Methods for Defining Evolutionary and Structural Relationships of External Solute-Binding Proteins.

[b] Residues in parentheses represent alternative possibilities at a particular position; triply underlined residues are fully conserved (invariant) residues in the multiple alignment. X, any residue.

[c] Two signature sequences are provided for clusters 3 and 6 (see text).

The periplasmic aliphatic-amino-acid-binding proteins are very closely related to each other, as revealed by the branch lengths in the phylogenetic tree, but they are relatively distant from the cytoplasmic AmiC repressor of the aliphatic amidase structural gene of *P. aeruginosa* (here designated AmiPae) (Fig. 4C). AmiPae is nevertheless strikingly similar to the binding proteins, particularly to the LIV-binding protein from *P. aeruginosa* (19 SD [Table 2]). The region of AmiC shown in Fig. 4B exhibits 32% sequence identity with LIVPae. Of the 88 residues depicted in Fig. 4B, 18 are fully conserved in the four members of this cluster.

AmiC undoubtedly binds aliphatic amides as inducers of expression of the aliphatic amidase structural gene (219), and these compounds are similar in structure to the aliphatic amino acids which bind to the periplasmic proteins of cluster 4. It had not previously been known to be homologous to other proteins in the data base (219).

TABLE 5. Mean similarity scores for the eight clusters of solute-binding proteins[a]

Cluster	Mean similarity score
Control[b]	0.23
1	0.45
2	0.52
3	0.65
4	0.77
5	0.52
6	0.90
7	0.42
8	0.40

[a] These values were taken from panels A of Fig. 1 to 6 and 8 and from Fig. 7.

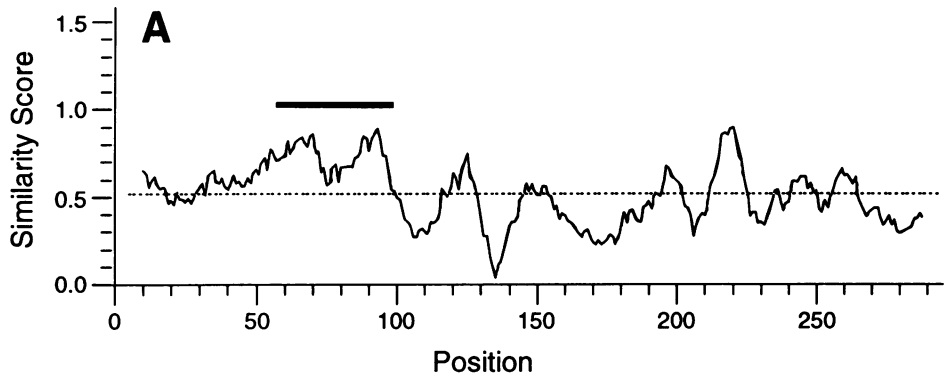
[b] The control value was obtained by plotting the similarity score for five nonhomologous binding proteins (AraEco, NocAtu, FecEco, SulEco, and CitSty [Table 1]) as a function of position in the multiple alignment and averaging the values obtained (Fig. 7). These proteins are all periplasmic binding proteins of similar sizes. When nonhomologous proteins of completely different functions are compared, lower mean comparison scores are generally obtained (unpublished results).

As with clusters 1 and 2, intercluster sequence similarities were observed between members of clusters 3 and 4. Most noteworthy were regions of alignment for HisSty with LIVEco and for LAOSty with both LeuEco and GlnEco (comparison scores of 6 to 7 SD for short regions exhibiting striking sequence similarity).

Cluster 5 Binding Proteins Specific for Peptides and Nickel

The peptide- and nickel-binding proteins of cluster 5 are the largest of the solute-binding proteins, with a size range of 493 to 543 residues (Tables 1 and 3). They are about twice the size of the cluster 3 proteins and the ChvE protein of cluster 2 (Table 3). Since they are more than 100 residues larger than any of the other binding proteins, they presumably contain one or more extra domains. As with the other binding-protein cluster similarity plots, the cluster 5 similarity plot revealed that the sequence similarity of these proteins is greatest in their N-terminal domains (Fig. 5A). An average similarity score of 0.52 (Table 5) shows that they exhibit a degree of similarity comparable to that observed for cluster 2 proteins. A portion of the multiple alignment of the peptide- and nickel-binding proteins of cluster 5 is shown in Fig. 5B, and the signature sequence for this family is presented in Table 4. The degree of sequence similarity revealed in this figure is somewhat lower than expected from the high comparison scores summarized in Table 2. This apparent discrepancy is because the cluster 5 binding proteins have substantially higher molecular weights than the other binding proteins do (Table 3). Larger proteins with comparable degrees of sequence similarity generally give higher comparison scores (see, for example, reference 158).

The relatively short branch lengths connecting several of the proteins in the phylogenetic tree (Fig. 5C) reflect the clear relatedness of these proteins. Four subclusters are apparent. The oligopeptide-binding proteins of *E. coli* and *Salmonella typhimurium* make up one subcluster, the oligopeptide- and dipeptide-binding proteins of *B. subtilis* make up the second subcluster, the dipeptide- and nickel-binding proteins of *E. coli* form the third subcluster, and the oli-



B

```
                  4 2 5 2 4 2 5 3 2 2 5 2 2 3 1 3 2 2 3 3 2 5 2 2 2 3 3 3 5 4 2 2 4 3 3 2 2 2
Rib Eco  (82) G T K I L L I N P T D S D A V G N A V K M A N Q A N I P V I T L D R Q A T K (119)
Rib Sty  (82) K T K I L L I N P T D S D A D G N A V K M A N Q A N I P V I T L D R Q A T K (119)
Chv Atu  (83) G V K V L V I A S I D G T T L S D V L K Q A G E Q G I K V I A Y D R L I R N (120)
Ara Eco  (57) G A K G F V I C T P D P K L G S A I V A K A R G Y D M K V I A V D D Q F V N (94)
Gal Eco  (59) G V K A L A I N L V D P A A A G T V I E K A R G Q N V P V V F F N K E P S R (96)
```

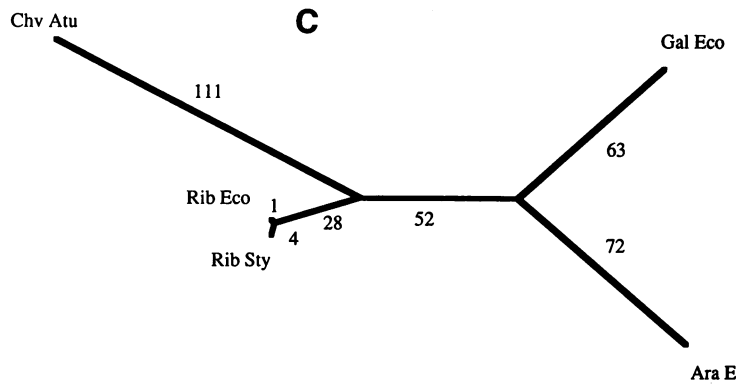


FIG. 2. Analyses of cluster 2 (monosaccharide-binding) proteins. Conventions are as described in the legend to Fig. 1, and the protein abbreviations are as presented in Table 1. (A) Similarity plot; (B) segment of the multiple alignment corresponding to the bar in panel A; (C) cluster 2 phylogenetic tree. Cluster 2 proteins are homologous to a large family of DNA-binding proteins as described previously (207).

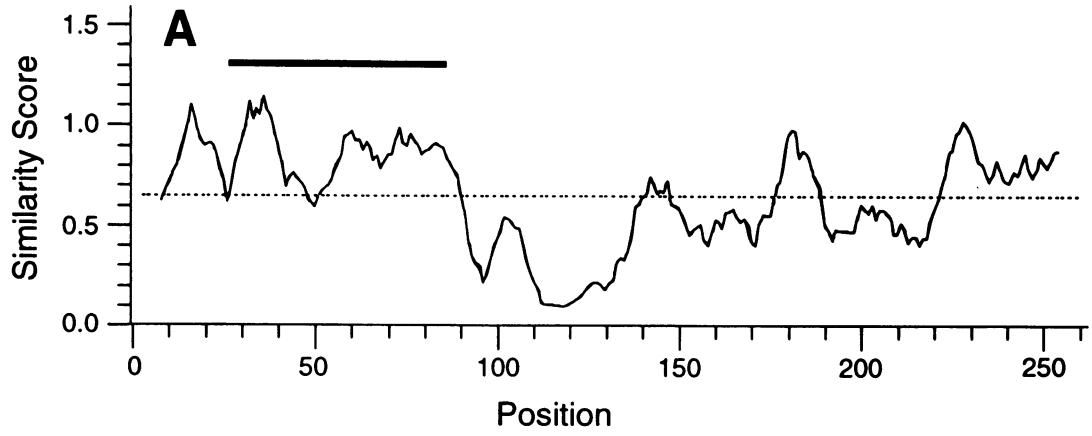
gopeptide-binding protein of *S. pneumoniae* represents the final subcluster. In view of its distinct binding specificity, it is surprising that the nickel-binding protein of *E. coli* (NikEco) is not more distant from the peptide-binding proteins than the latter are from each other. This fact suggests that nickel is accommodated in a fashion which does not require extensive modification of the active site (see below).

One additional sequenced protein (of unknown function) proved to be homologous to the peptide-binding proteins of cluster 5. This protein, designated XP55, is a major secreted protein of 55 kDa from *Streptomyces lividans* (24). It was found to be 25% identical with the oligopeptide-binding protein of *S. typhimurium* (OppSty) over an overlapping segment of 312 amino acyl residues (Table 2). It is the same size as the other cluster 5 proteins (Table 1). The comparison score of 24 SD establishes that this protein is homologous with the peptide-binding proteins depicted in Fig. 5. Its similarity to the cluster 5 proteins suggests that it might be a peptide-binding protein.

Cluster 6 Binding Proteins Specific for Inorganic Polyanions

The sulfate- and thiosulfate-binding proteins of enteric bacteria and cyanobacteria are all closely related (comparison scores of 36 to 80 SD [Table 2]). They occur in a size range of 310 to 350 residues (Table 3). Figure 6A shows the similarity plot, Fig. 6B shows a segment of the multiple alignment, and Fig. 6C shows the tree. In the 39-residue multiple alignment shown in Fig. 6B, 21 residues (54%) are fully conserved. Two signature sequences for these proteins are presented in Table 4.

Because the phosphate-binding protein of *E. coli* gave low comparison scores with these proteins (Table 2), it could not be established as a member of the cluster and was omitted from the analyses reported in Fig. 6. It should be noted, however, that it is in the same size range as the sulfate-binding protein (Table 3), it binds an inorganic anion, and it exhibits a comparison score of 6 SD with the sulfate-binding protein of *E. coli* (Table 2). When the phosphate- and sulfate-binding proteins of *E. coli* were examined for simi-



B

```
                  6 6 5 4 6 6 2 4 3 3 4 5 3 3 3 1 3 4 2 3 4 2 3 2 3 4 3 3 4 6 4 6 2 3 4 3 2 6 4 4 4 3 2 2 3 5 5 3 3 6 3 3 2 5 3 6 5
LAO Sty  (50) G F D I D L G N E M C K R M Q V K C T W V A S D F D A L I P S L K A K K I D A I I S S L S I T D K R Q Q E I A F S (106)
His Sty  (50) G F D I D L A K E L C K R I N T Q C T F V E N P L D A L I P S L K A K K I D A I M S S L S I T E K R Q Q E I A F T (106)
Gln Eco  (48) G F D V D L W A A I A K E L K L D Y E L K P M D F S G I I P A L Q T K N V D L A L A G I T I T D E R K K A I D F S (104)
Occ Atu  (47) G F E I D L A N A L C E K M K A K C Q I V A Q N W D G I M P S L T G K K Y D A I M A A M S V T P K R Q E V I G F S (103)
Noc Atu  (53) G F D I D L G N D L C K R M N I E C K F V E Q A W V G I I P S L T A G R Y D A I M A A M G I Q P A R E K V I A F S (109)
Baa Ngo  (61) G F D V D L M N A M A K A G N F K I E F K H Q P W D S L F P A L N N G D A D I V M S G V T I T D D R K Q S M D F S (117)
```

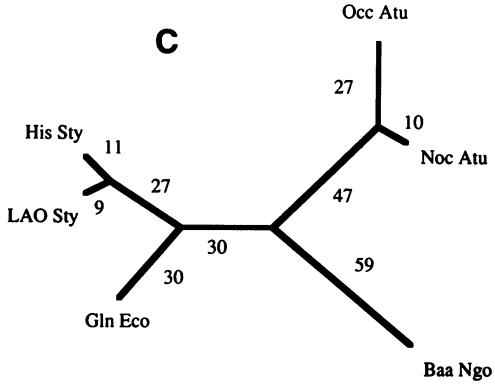
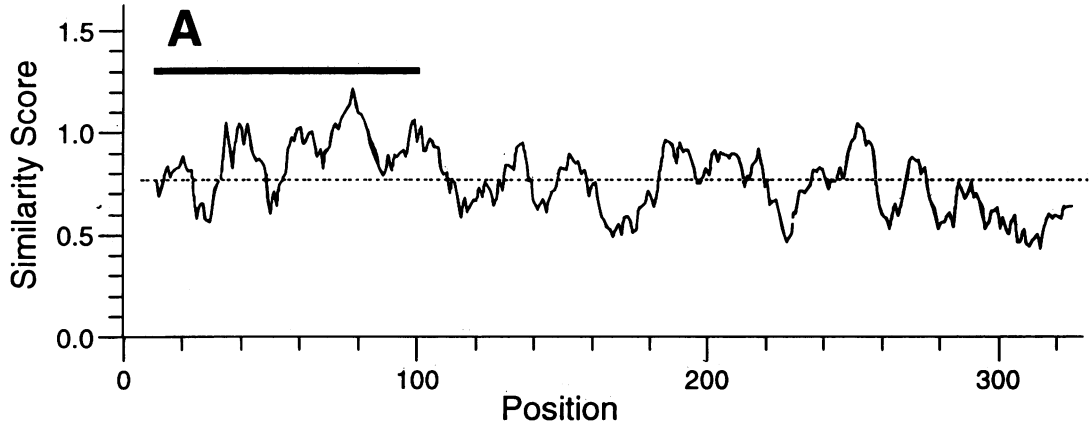


FIG. 3. Analyses of cluster 3 (polar-amino-acid- and opine-binding) proteins. Protein abbreviations are as presented in Table 1, and conventions used are as described in the legend to Figure 1. (A) Similarity plot; (B) segment of the multiple alignment corresponding to the bar in panel A; (C) cluster 3 phylogenetic tree.

larity by using a similarity plot like that shown for the homologous protein members of the various clusters, an average similarity score of 0.39 was obtained. Although this value is lower than those observed for the established binding-protein clusters (Table 5), it is almost as high as the value obtained for cluster 8 proteins, and it is very substantially above that obtained for the intercluster comparison of binding proteins used to obtain a control value (average similarity score of 0.23 [Table 5; see also Fig. 7]).

Cluster 7 Binding Proteins Specific for Organic Polyanions

Cluster 7 consists of only two proteins, one specific for dicarboxylates and the other specific for tricarboxylates. They consist of 303 and 333 residues, respectively (Tables 1 and 3). They exhibit a fairly low comparison score (7 SD [Table 2]) but nevertheless exhibit regions of striking sequence similarity (Fig. 7). In Fig. 7, the light line (top) shows the similarity plot for DcaRca versus CitSty and the heavy line (bottom) shows the similarity plot for five nonhomologous binding proteins from five different clusters. These five binding proteins are all about the same size (see the legend to Fig. 7). The average similarity score for this intercluster comparison was 0.23 (bottom dashed line in Fig. 7; Table 5). The average similarity score obtained when DcaRca and CitSty were compared (top dashed line in Fig. 7) was 0.42, a value which is substantially larger than that obtained for the control proteins and which is intermediate between the values obtained for cluster 1 proteins and cluster 8 proteins (Table 5). Throughout most of the plot, the degree of similarity was greater for DcaRca and CitSty than for the control proteins. The region of greatest similarity was found to be in the N-terminal domain, as indicated by the bar in Fig. 7. This region of the binary alignment corresponds to



B

```
                    2 4 3 2 4 3 2 3 2 2 2 2 1 4 4 2 2 4 2 2 2 3 4 2 2 4 4 2 2 4 1 2 3 2 2 3 1 3 4 3 3 3 4 4 3

Leu  Eco  (35)  S G P I A Q W G I M E F N G A E Q A I K D I N A K G G I K G D K L V G V E Y D D A C D P K
LIV  Eco  (12)  S G P V A Q Y G D Q E F T G A E Q A V A D I N A K G G I K G N K L Q I V K Y D D A C D P K
LIV  Pae  (39)  T G P V A Q Y G D M Q R A G A L M A I E Q I N K A G G V N G A Q L E G V I Y D D A C D P K
AmiC Pae  (17)  T G V T A D I E R S H A Y G A L L A V E Q L N R E G G V G G R P I E T L S Q D P G   D P D

                    3 3 3 3 3 4 3 3 2 3 3 3   4 2 3 2 3 2 4 3 2 3 4 3 3 3 3 3 3 2 3 3 3 4 3 3 3 2 4 3 3 2 4

Leu  Eco          Q A V A V A N K I V N D   G I K Y V I G H L C S S S T Q P A S D I Y E D E G I L M I S P  (122)
LIV  Eco          Q A V A V A N K V V N D   G I K Y V I G H L C S S S T Q P A S D I Y E D E G I L M I T P  (99)
LIV  Pae          Q A V A V A N K V V N D   G V K F V V G H V C S S S T Q P A T D I Y E D E G V L M I T P  (126)
AmiC Pae          R Y R L C A E D F I R N R G V R F L V G C Y M S H T R K A V M P V V E R A D A L L C Y P  (105)
```

C

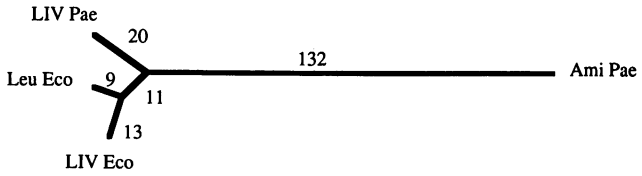


FIG. 4. Analyses of cluster 4 (aliphatic, hydrophobic-amino-acid-binding) proteins. Protein abbreviations are as presented in Table 1, and conventions used are as described in the legend to Fig. 1. (A) Similarity plot; (B) segment of the multiple alignment corresponding to the bar in panel A; (C) cluster 4 phylogenetic tree. The AmiC (AmiPae) protein is a cytoplasmic transcriptional repressor which regulates induced synthesis of the aliphatic amidase of *P. aeruginosa* (219).

the following alignment:

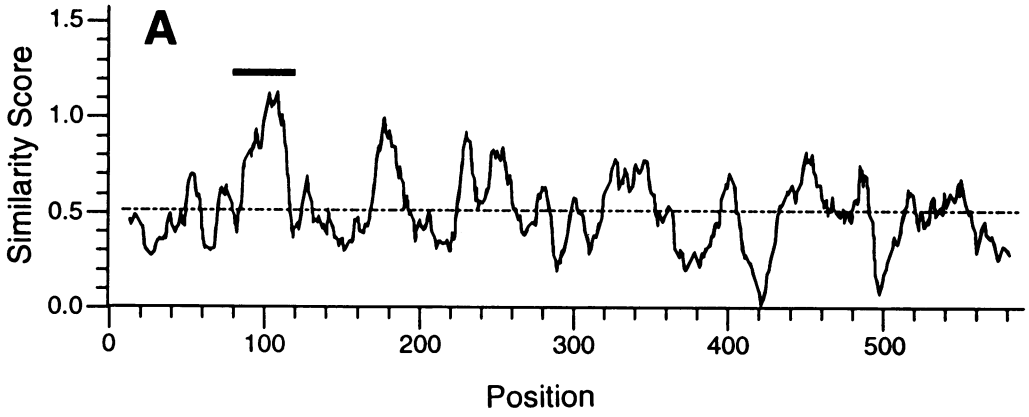
```
                                                                  106
DcaRca:  L G A V Q M L A P S L A K F G P L G V Q D F E
         . . .   : :   : . : : :   . : : : . : . .
CitSty:  F S G G S L L N L S Q G K F G R Y G V D D V R
                                                    93
```

These two segments of 23 overlapping residues show 35% sequence identity and 74% sequence similarity. The proposed signature sequence for these two cluster 7 proteins is presented in Table 4. The comparison score of 7 SD, the similar solute-binding specificities of these two proteins, their similar sizes, and the fact that the region of greatest similarity between them is found within their N-terminal domains all suggest that these two proteins are homologous.

As with clusters 1 and 2 and clusters 3 and 4, sequence similarities were recorded between members of clusters 6 and 7. For example, CitSty and PhoEco gave a comparison score of 6 SD for a short overlapping segment. Such limited degrees of sequence similarity probably did not arise by chance, but they could have arisen by either convergent or divergent evolutionary processes.

Cluster 8 Binding Proteins Specific for Iron Complexes (and Possibly for Vitamin B_{12})

Cluster 8 binding proteins, specific for iron complexes, exist in a narrow size range (296 to 322 residues) (Tables 1 and 3). Köster et al. (98) have shown that periplasmic binding proteins of *E. coli* specific for iron complexes and vitamin B_{12} display significant sequence similarity and have published the complete multiple alignment of available cluster 8 sequences. The similarity plot (Fig. 8A), a portion of



B

```
                    6 4 3 7 5 4 6 4 3 5 5 7 4 6 5 4 7 6 3 4 7 3 3 4 7 5     5 4 6 2 5 4 4 5 2 4 4
Opp Eco  (84) P G V A E S W D   N K D A K V W T F H L R K D A K W     S D G T P V T A Q D F  (119)
Opp Sty  (83) P G V A E K W E   N K D F K V W T F H L R E N A K W     S D G T P V T A H D F  (118)
Dpp Bsu  (89) P A M A E K W S V S K D N K T Y T F T I R E N A K W     T N G D P V T A G D F  (125)
Opp Bsu  (85) E G M A S K I E T S K D G K T Y T F T I R D G V K W     S N G D P V T A Q D F  (121)
Dpp Eco  (76) P G L A E K W E V S E D G K T Y T F H L R K G V K R     H D N K E F K P T R E  (112)
Nik Eco  (70) P W L A K S W T H S E D G K T W T F T L R D D V K F     S N G E P F D A E A A  (106)
Opp Spn  (80) P A V A E D W E V S K D G L T Y T Y K I R K G V K W F T S D G E E Y A E V T A  (118)
```

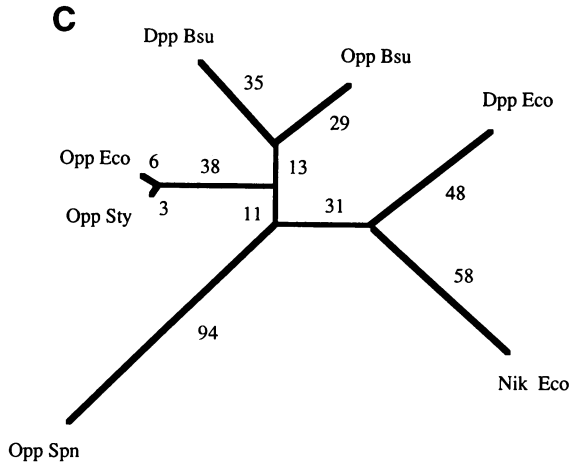
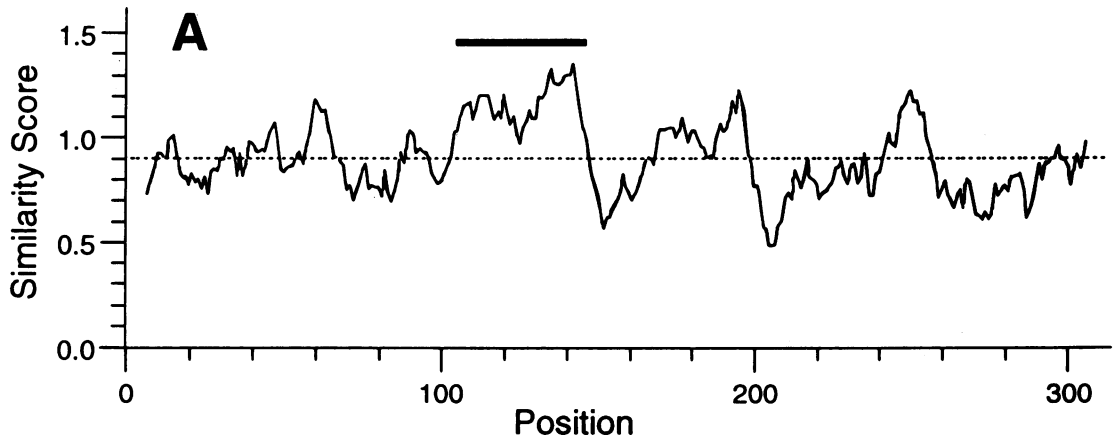


FIG. 5. Analyses of cluster 5 (oligopeptide-, dipeptide-, and nickel-binding) proteins. Protein abbreviations are as presented in Table 1, and conventions are as described in the legend to Fig. 1. (A) Similarity plot; (B) segment of the multiple alignment corresponding to the bar in panel A; (C) cluster 5 phylogenetic tree.

the multiple alignment (Fig. 8B), and the phylogenetic tree (Fig. 8C) for the four iron-complex-binding proteins, three from *E. coli* and one from *Vibrio anguillarum*, provide a basis for evaluating the sequence similarities of these proteins. The first three proteins are the FepB protein (specific for ferrienterobactin), the FecB protein (specific for iron dicitrate), and the FhuD protein (specific for iron hydroxamates such as ferrichrome and aerobactin). The four iron-complex-binding proteins which make up the cluster 8 family show the lowest average similarity score of the proteins characterized in Fig. 1 to 8 (Table 5). The signature sequence for this family (Table 4) is specific for the four members of cluster 8. It is derived from the portion of the multiple alignment shown in Fig. 8B, which is the most strongly conserved region within these proteins (Fig. 8A). As with the other clusters, this region is present within the N-terminal domains of these proteins. The phylogenetic tree (Fig. 8C) reveals that the four protein members of cluster 8 are about equidistant from each other.

The vitamin B_{12}-binding protein (BtuE or BtuEco) exhibits a poor comparison score with and is much smaller than the other cluster 8 proteins (Tables 1 to 3). Interestingly, the BtuE protein is homologous to human glutathione peroxidase (6). These two proteins exhibit 41% sequence identity over an overlapping region of 156 residues (unpublished results). For this reason it is particularly important to emphasize that homology of this protein with the other members of cluster 8 is not established. The periplasmic iron-binding proteins of the *Serratia marcescens* and *Neisseria* iron transport systems (cluster 1) show an insignificant degree of sequence similarity to the proteins of cluster 8 (unpublished results).



B

```
                    4 4 4 4 2 4 2 4 4 3 4 4 2 3 3 3 4 2 2 3 3 3 4 4 4 2 4 4 3 4 4 3 3 3 4 3 4 4 4
Sul Eco  (123) G N P K Q I H D W N D L I K P G V S V I T P N P K S S G G A R W N Y L A A W G  (161)
Sul Sty  (105) G N P K Q I T D W N D L I K P G V S V I T P N P K S S G G A R W N Y L A A W G  (143)
Tsu Eco  (129) G N P K N I H D W N D L V R S D V K L I F P N P K T S G N A R Y T Y L A A W G  (167)
Sul Ssp  (143) G N P K G I K D W T D L T K P G V R I V T A N P K P S G G A R W N F L G A W G  (181)
```

C

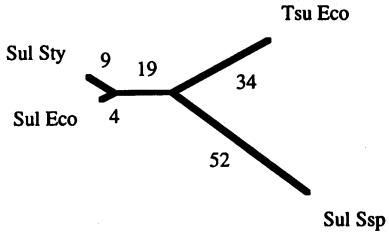


FIG. 6. Analyses of the cluster 6 (inorganic anion-binding) proteins. Protein abbreviations are as presented in Table 1, and conventions are as described in the legend to Fig. 1. (A) Similarity plot; (B) segment of the multiple alignment corresponding to the bar in panel A; (C) cluster 6 phylogenetic tree. The phosphate-binding protein of *E. coli* (Table 1) is excluded from the figure because of the low comparison score obtained between this protein and other cluster 6 proteins (Table 2).

BIOCHEMICAL PROPERTIES OF PERIPLASMIC BINDING PROTEINS

Biochemical and genetic experiments have established the existence of several ABC-type uptake transporters for which sequence data are not yet available. Among these transport systems are four permeases of *E. coli*: the cystine permease (18, 109), a glutamate-aspartate permease (104), a xylose permease (177), and a thiamine permease (90, 137). In all four cases, the presence of solute-binding proteins has been established and the properties of the purified proteins have been described. Some of the biochemical characteristics of these proteins are summarized and compared with those of other representative binding proteins in Table 6. The reported molecular weights of the cystine- and the glutamate-aspartate-binding proteins (28,000 to 30,000) are similar to those specific for other hydrophilic amino acids (Table 3). These proteins may therefore prove to fall into cluster 3. The molecular weight reported for the xylose-binding protein (37,000) is in between those for the two sugar-binding protein families, but, on the basis of the nature of the sugar bound, one would expect this protein to fall into cluster 2. The molecular weight of the thiamine-binding protein is unknown.

The dissociation constants (K_d values) of these and other representative well-characterized binding proteins for their solutes are summarized in Table 6 and are compared with the apparent affinity constants (K_m values) of the corresponding transport systems for the same solutes. It can be seen that the K_d values range from 5 nM for vitamin B_{12} and 10 nM for cystine to 1 to 2 μM for maltose, galactose, glutamate-aspartate, and citrate. Reported K_m values are generally close to the K_d values, and, with only one exception, the former values are always within an order of magnitude of the latter values. The one exception is the sulfate-binding protein, which is reported to have a K_d for sulfate of 20 nM (146) whereas the K_m of the corresponding transport system for

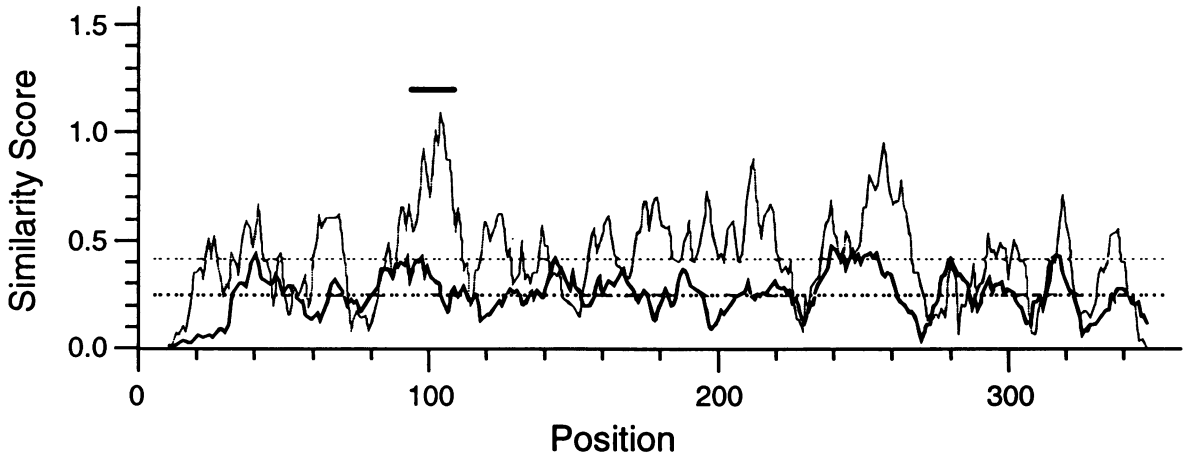


FIG. 7. Similarity plot for the two cluster 7 binding proteins, the dicarboxylate-binding protein of *Rhodobacter capsulatus* (DcaRca) and the tricarboxylate-binding protein of *Salmonella typhimurium* (CitSty). This similarity plot is shown by a light line (top), and the control similarity plot is shown by a heavy line (bottom). Throughout most of their lengths, DcaRca and CitSty exhibit a greater degree of similarity than do the control proteins. The latter proteins include AraEco (cluster 2), NocAtu (cluster 3), SulEco (cluster 6), CitSty (cluster 7), and FecEco (cluster 8) (Table 1). Average similarity scores for each of these two groups of aligned sequences are indicated by the dashed lines, and the values are tabulated in Table 5.

sulfate is 36 μM (45). Confirmatory experiments should be conducted to determine whether this 2,000-fold difference is real or artifactual. The fact that the K_d and K_m values are generally so similar suggests that the binding protein is largely responsible for the apparent affinity of the transport system for its substrate(s). This suggestion is in agreement with the fact that the binding specificities of these proteins correlate closely with the substrate specificities of the transport systems. Biochemical properties of the well-characterized periplasmic binding proteins of enteric bacteria have been comprehensively reviewed by Furlong (55), and the interested reader is referred to this article for further details.

EVOLUTIONARY CONSTRAINTS IMPOSED BY STRUCTURE AND FUNCTION

The analyses reported in this review form the first comprehensive sequence comparison study concerned with the structural and evolutionary relatedness of the gram-negative bacterial periplasmic and gram-positive bacterial membrane-bound solute-binding proteins. These proteins provide a unified receptor function by interacting with transmembrane transducer proteins which mediate chemotaxis and transport and sometimes initiate pathways of sensory transduction. X-ray crystallographic analyses of several of these proteins, reported primarily from the laboratory of F. Quiocho, have shown that although the proteins have diverged very substantially with respect to specific structural features, they nevertheless share many essential three-dimensional structural motifs (154, 209). The results of the sequence comparisons reported here establish that certain of these proteins are homologous and provide compelling evidence for the evolutionary relatedness of several others. Thus, all of the proteins within a particular cluster and depicted on each of the phylogenetic trees (Fig. 1 to 6 and 8) clearly share an evolutionary origin. It is interesting that two clusters each were found for sugars, amino acids, and anions and that the nature of the substrates bound in each of the subclusters differed (oligosaccharides versus monosaccharides for the two sugar clusters, hydrophilic versus hydrophobic amino acids for the two amino acid clusters, and inorganic versus organic anions for the two clusters of anions). Within each cluster of proteins, uniform molecular size was observed with only a few exceptions (Table 3).

The observation that the relatedness of the binding proteins generally follows their binding specificities was unexpected. The solute-binding site, localized to the crevice between the two domains of each of these proteins, represents only a small part of the entire structure. Despite this, it seems that the requirement for solute binding imposes significant structural constraints on the protein. This implies, first, that a major part of the protein plays a role, primary or secondary, in solute binding and, second, that serial mutations which alter the binding site can allow accommodation of solutes of a related nature much more readily than solutes which exhibit greater structural divergence. Although these conclusions seem reasonable in retrospect, they were not apparent at the onset of our studies.

The generalization noted above, that binding specificity correlates with the nature of the ligand bound, was not always observed. Thus, the iron-binding proteins of *Serratia* and *Neisseria* species are apparently homologous to the carbohydrate-binding proteins of cluster 1, the nickel-binding protein of *E. coli* is homologous to the peptide-binding proteins of cluster 5, and the vitamin B_{12}-binding protein may be homologous to the iron-complex-binding proteins of cluster 8. The structural bases for these observations must await elucidation of the three-dimensional structures of the homologous proteins. However, the high-affinity binding of inorganic ionic species, both anionic and cationic, within the domains of enzymes which do not act on these species is not at all unusual, and a Ca^{2+}-binding site in the galactose-binding protein of *E. coli* has been noted (130, 208).

Periplasmic solute-binding proteins from evolutionarily divergent organisms which are specific for the same solute are, in general, much more closely related to each other than are proteins specific for different solutes from the same organism. This observation suggests that the proteins specific for the various solutes arose by gene duplication and divergence events which preceded evolutionary divergence of the species. Alternatively, horizontal transmission of genetic material during the early evolution of the species

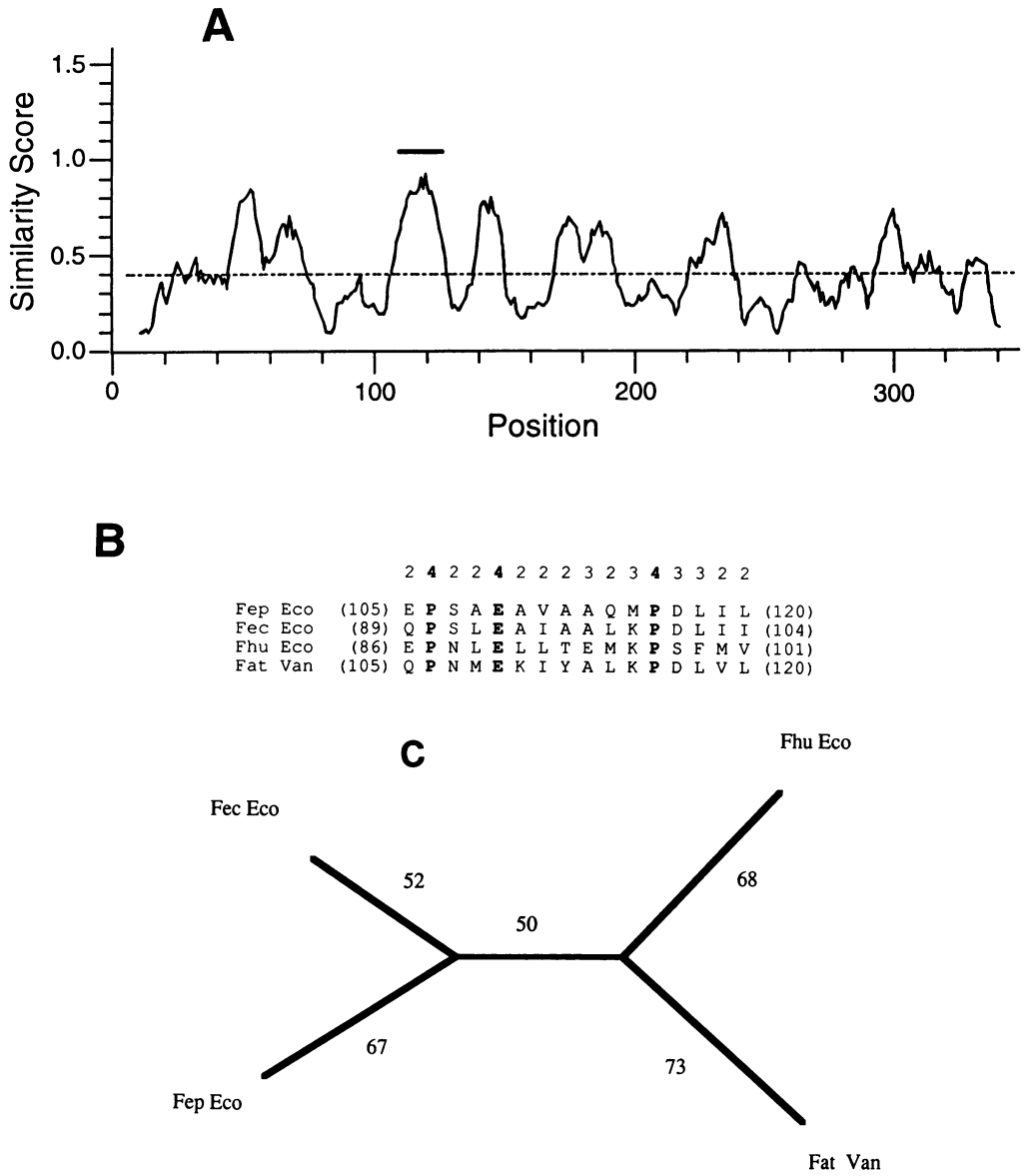


FIG. 8. Analyses of cluster 8 (iron-complex-binding) proteins. The abbreviations and conventions are as presented in Table 1 and the legend to Fig. 1, respectively. (A) Similarity plot; (B) segment of the multiple alignment corresponding to the bar in panel A; (C) phylogenetic tree. The complete multiple alignment of these four proteins together with the vitamin B_{12}-binding protein of *E. coli* (BtuEco) has been published (98).

may have occurred. The same conclusion has recently been reached regarding the permease proteins of the bacterial phosphotransferase system (110, 156). Assuming the former explanation (vertical transmission of genetic material) to be correct, the many solute-binding proteins with a diversity of substrate specificities must have existed as distinct species prior to divergence of the major bacterial kingdoms (i.e., gram-positive bacteria from gram-negative bacteria and cyanobacteria from purple bacteria). Thus, these proteins may have existed more than two billion years ago in forms similar to those found today.

Three-dimensional analyses of numerous periplasmic binding proteins of gram-negative bacteria have revealed that they are bilobed, two-domain structures which probably arose by a duplication event from a primordial gene encoding a simple one-domain protein. Each of the two domains in the binding proteins consists largely of β-structure while the connecting loops are rich in α-helix (154). In the study reported here, we have unexpectedly found that within each cluster of homologous binding proteins, the sequences of the N-terminal domains are conserved to a much greater degree than are the C-terminal domains. Although the significance of this intriguing observation is not clear, it could suggest that the functional and structural constraints imposed on the N-terminal domains are more rigid and that these domains have therefore undergone evolutionary divergence more slowly than the C-terminal domains. It further suggests that the two domains are nonequivalent. The conclusion of nonequivalence has been reached for the other constituents of ABC-type transporters (181a). It is possible that the

TABLE 6. Biochemically determined properties of representative binding proteins from enteric bacteria

Solute-specific binding protein[a]	Mol mass (kDa) (approx)	K_d (μM)	K_m (μM)	References
Sugars				
Maltose	41	1	1	95, 197
α-Glycerol-P	45	0.2	2	13, 176
Xylose	37	0.6	5	5
Ribose	30	0.1	0.3	217
Galactose	32	1	0.5	12, 20, 166
Amino acids				
LAO	28	0.2	0.5	162,
Cystine	28	0.01	0.1	18, 142
Glutamine	26	0.2	0.6	213, 214
Glutamate-aspartate	30	1	0.5	16, 218
LIV or leucine	37	0.4	0.3	56, 57, 150
Anions				
Citrate	28	2	3	196
Phosphate	34	0.8	0.2	114, 126, 163, 195
Sulfate	35	0.02	36	45, 146
Vitamins				
B_{12}	22	0.005	0.01	22, 199
Thiamine	?	0.1	0.8	90, 137

[a] The organism from which the binding proteins were isolated was always *E. coli* or *S. typhimurium*.

primordial protein of about the size of a current binding-protein domain corresponded more closely in structure and function to the N-terminal domain, which has consequently evolved more slowly than the C-terminal domain. Since some binding proteins are more than twice as large as others (Table 3), we expect that the larger ones, particularly the cluster 5 proteins, contain additional domains. Analyses of primordial sequences, determined from present-day protein sequences, may provide information on these possibilities.

RELATIONSHIPS BETWEEN EXTERNAL RECEPTORS AND CYTOPLASMIC REPRESSORS

It has recently been demonstrated (122, 123, 207, 212) that the bacterial periplasmic sugar-binding proteins of cluster 2 are structurally and evolutionarily related to a large class of bacterial cytoplasmic helix-turn-helix-type DNA-binding proteins, many of which are repressors which bind sugar inducers. Over a dozen transcriptional regulatory proteins are members of this subfamily, all of which are homologous to the cluster 2 binding proteins. This fact greatly increases the size of the cluster 2 family of proteins. The phylogenetic tree of the DNA-binding-protein family, showing the positions of these proteins relative to each other as well as to the ribose-binding proteins of *E. coli* and *Salmonella typhimurium*, has been published (207). The structures of the homologous DNA-binding proteins have not yet been elucidated in three dimensions, but the demonstration that they are homologous to the binding proteins suggests that their sugar-binding domains exhibit the bilobed tertiary structures characteristic of the periplasmic binding proteins (154).

The two major classes of functionally divergent cluster 2 proteins (receptors and repressors) differ with respect to their subcellular localizations as well as their biological functions. The subcellular localization of the periplasmic binding proteins (presumed to be determined by an N-terminal signal sequence [155, 172, 174, 216]), as well as the DNA-binding characteristics of the repressor proteins (presumed to be determined by an N-terminal helix-turn-helix motif [207, 212]), is not shared by these two classes of proteins. These differences presumably arose by gene splicing and fusion events during evolutionary history. Despite these important structural differences, the proteins share the characteristic that binding of the small molecular ligand (usually a sugar molecule) controls the conformation of the protein and consequently its intermolecular interactions. Thus, sugar binding increases the affinity of the periplasmic receptor proteins for the transmembrane transducer proteins but decreases the affinity of the cytoplasmic repressor proteins for the operator regions of their target operons. This observation suggests a unified function for solute binding and a possible explanation for the common origin of receptor and regulatory proteins. A common evolutionary origin for receptor, regulatory, and transport proteins had been postulated previously (170).

A second example of homology between a periplasmic binding protein and a cytoplasmic transcriptional regulatory protein was noted in the present study. AmiC, the repressor of the structural gene encoding the aliphatic amidase of *P. aeruginosa*, here designated AmiPae, is clearly homologous to the hydrophobic aliphatic-amino-acid-binding proteins of cluster 4 (Fig. 4). Although the detailed mode of action of AmiC is not known, the homology noted here provides significant clues about its structure, the nature of the inducers which bind to AmiC, and the nature of the conformational change which regulates its affinity for the operator of the *ami* operon.

EVOLUTION OF COMPLEX, MULTICOMPONENT TRANSPORT SYSTEMS

In the present study, we have analyzed only one component of the bacterial ABC-type transport systems, namely the periplasmic solute-binding proteins. Two other sets of protein constituents of these systems are available for analysis: the integral transmembrane constituents or domains and the ATP-binding constituents or domains (3, 55, 75, 78). If most or all of the ABC-type transporters evolved as a unit (i.e., a genetic transcriptional unit) from a common ancestral transport system (or a primordial operon), one would expect that the relative rates of evolutionary divergence would have been about the same for the three dissimilar constituents of the systems, even if the absolute rates of divergence were markedly different. Our preliminary results have suggested that, for at least some of these systems, the periplasmic binding proteins diverged from the common ancestral protein most rapidly, that the ATP-binding protein constituents of the ABC-type transporters diverged least rapidly, and that the integral transmembrane constituents of these systems evolved at an intermediate rate (unpublished results). Further, it appears that the integral proteins of many of the bacterial ABC-type transporters fall into clusters similar to those noted here for the binding proteins (34a; unpublished results). If this proves to be true for the majority of the bacterial ABC-type transporters, it may be possible to confirm the cluster-type grouping demonstrated here for the binding proteins for the transport systems as a whole, and it may also be possible to provide evidence for the interrelatedness of these clusters. Moreover, a positive correlation of relative evolutionary rate would strongly suggest that these systems did, in fact, evolve as complete transport systems

rather than as individual proteins which came together late in the evolutionary process to form the transport systems. It should be noted, however, that data discrepant with this simple possibility have been published for one subclass of ABC-type transporters (159).

Almost all large superfamilies of proteins have members in both the eukaryotic and prokaryotic kingdoms. For example, this observation has been made for the ATP-binding constituents of the ABC-type transporters (159); for a major superfamily of uniport, symport, and antiport facilitators (119); for a large group of amino acid facilitators (157); for P-type ATPases (49a); and for a variety of enzymes (201). In a few instances, members of large protein superfamilies have been restricted to a single kingdom. Thus, the permeases of the bacterial phosphotransferase system (156) which have bacterial non-phosphotransferase system homologs (160) have no established counterpart in eukaryotes. Conversely, the voltage-gated ion channel proteins of eukaryotes have no known prokaryotic counterpart (171). The periplasmic binding proteins appear to form another prokaryote-specific family. With the sole exception of the vitamin B_{12}-binding protein, which is homologous to the human glutathione peroxidase (6), no eukaryotic homologs of these proteins have been detected. If the BtuE protein proves to have a structure and/or function other than that of a typical periplasmic binding protein, a possibility we consider to be quite likely, it follows that there are no eukaryotic homologs of the typical extracellular solute-binding proteins of bacteria. This observation is particularly interesting since members of this proposed prokaryotic superfamily include periplasmic binding proteins of gram-negative bacteria, cell surface lipoprotein receptors of gram-positive bacteria, and cytoplasmic DNA-binding proteins of both gram-positive and gram-negative bacteria (207, 212; also see above). It may be that eukaryotic members of this family will be found. However, an as-yet-unknown basis for this restriction may exist. Dozens of eukaryotic transporters of the ABC-type family are known, but none of them apparently utilizes an extracytoplasmic solute-binding (lipo)protein (75, 78). There is no a priori reason why solute-binding lipoprotein-dependent ABC-type uptake systems analogous to those found in gram-positive bacteria could not be functional in animal, plant, and fungal cells as well as in unicellular eukaryotes. Further work is required to resolve this intriguing enigma.

ARE THERE ADDITIONAL BINDING-PROTEIN FAMILIES?

We have noted that a few sequenced bacterial extracytoplasmic solute-binding proteins do not fall within the eight clusters of proteins described in this report. These proteins include the glycine-betaine-binding proteins, ProX, of *E. coli* (67, 124) and of *Salmonella typhimurium* (144, 190); the phosphonate-binding protein, PhnD, of *E. coli* (30, 116, 211); the polyamine-binding protein of *E. coli*, PotD (58); the putative copper-binding protein of *Pseudomonas stutzeri*, NosD (227); the mercuric-ion-binding proteins, MerP, of gram-negative bacteria (139); and the p37 protein (of unknown specificity) of *M. hyorhinis* (46). Additionally, we have summarized biochemical evidence for the existence of binding proteins specific for cystine, glutamate-aspartate, xylose, and thiamine (55). Other binding proteins will undoubtedly be discovered, particularly as a result of prokaryotic-genome sequencing and possibly as a result of eukaryotic-genome sequencing as well. It is therefore likely that additional families of bacterial periplasmic binding proteins and their homologs will emerge as the genes encoding these proteins are sequenced.

ACKNOWLEDGMENTS

We thank Jorge H. Crosa, Oregon Health Science University; William W. Kay, University of Victoria; David J. Kelley, University of Sheffield; Abraham Sonnenshein, Tufts University School of Medicine; and Long-Fe Wu, Institut National de Science Appliqués, Villeurebanne, France, for providing sequence data and information prior to publication. We are grateful to Gerardo M. Pao, Aiala Reizer, and Jonathan Reizer for valuable assistance with the computer analyses reported. Aiala and Jonathan Reizer generously screened the signature sequences against the data bases and generated the similarity plots presented in this paper. Mary Beth Hiller and Melinda Marcus provided expert word-processing assistance.

This work was supported by Public Health Service grants 5RO1AI21702 and 2RO1AI14176 from the National Institute of Allergy and Infectious Diseases.

ADDENDUM IN PROOF

Since the completion of this review several interesting advances in periplasmic solute-binding receptor research have been made. A bacterial periplasmic receptor homolog with catalytic activity has been identified (R. Tam and M. H. Saier, Jr., Res. Microbiol., in press). This enzyme is cyclohexadienyl dehydratase of *P. aeruginosa* (G. Zhao, T. Xia, R. S. Fischer, and R. A. Jensen, J. Biol. Chem. **267**:2487–2493, 1992). This periplasmic enzyme is homologous to periplasmic receptors for polar amino acids (cluster 3). The structure of its substrate-binding domain must be closely related to those of the histidine- and LAO-binding proteins. The latter protein has recently been resolved in three dimensions at high resolution by X-ray crystallography, both with and without the ligand, lysine (B.-H. Oh, J. Pandit, C.-H. Kang, K. Nikaido, S. Gokcen, G. F.-L. Ames, and S.-H. Kim, J. Biol. Chem., in press). Based on analogy with the LAO-binding protein, the known active-site threonyl residue in cyclohexadienyl dehydratase of *P. aeruginosa* was inferred to be localized to the cleft region where the receptors bind their solutes (Tam and Saier, in press).

The genes for the ABC-type molybdenum transport system from *Azotobacter vinelandii* have recently been sequenced (F. Luque, A. Mitchenall, M. Chapman, R. Christine, and R. N. Pau, Mol. Microbiol. **7**:447–459, 1993). Four open reading frames were identified, ORFs 1 to 4. The genetic structure of the operon and the predicted products of these four ORFs were suggestive of an ABC-type transporter. ORFs 3 and 4 were shown to be homologous to ChlD and ChlJ (partially sequenced) from *E. coli*. The latter two proteins are parts of a putative molybdenum transporter. ORF 3 is an integral membrane protein, while ORF 4 is the ATP-binding protein. The authors suggested that ORF 2 is a transmembrane protein while ORF 1 is a periplasmic binding protein. However, ORF 2 exhibits the hydropathy characteristics of a hydrophilic protein, except for a putative, cleavable N-terminal signal sequence. Further, ORF 1 lacks a signal sequence and exhibits homology with molybdenum-pterin-binding proteins (Mop I to Mop III) from *Clostridium pasteurianum* (Luque et al., Mol. Microbiol., 1993). We suggest that the functional assignments of these proteins by Luque et al. (Mol. Microbiol., 1993) is incorrect and that ORF 2 is the periplasmic binding protein while ORF 1 performs a function in the cytoplasm. If this is the case, then a single transmembrane protein is present in this permease, presumably forming a homodimeric species. Such a situation has been noted for other ABC-type permeases (159).

An operon encoding a novel putrescine-specific transport system of the ABC type has been sequenced (R. Pistocchi, K. Kashiwagi, S. Miyamoto, E. Nukui, Y. Sadakata, H. Kobayashi, and K. Igarashi, J. Biol. Chem. **268:**146–152, 1993). This operon maps at 19 min on the *E. coli* chromosome. It consists of four ORFs, *potF*, *-G*, *-H*, and *-I*. The PotF protein was shown to be a putrescine-specific periplasmic binding protein. It was homologous (35% identity) to the spermidine- (putrescine-) binding protein PotD, sequenced previously, which is present within a distinct ABC-type transport operon which maps at 15 min on the *E. coli* chromosome (58).

REFERENCES

1. **Abouhamad, W. N., M. Manson, M. M. Gibson, and C. F. Higgins.** 1991. Peptide transport and chemotaxis in *Escherichia coli* and *Salmonella typhimurium*: characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein. Mol. Microbiol. **5:**1035–1047.
2. **Actis, L. A., M. E. Tolmasky, D. H. Farrell, and J. H. Crosa.** 1988. Genetic and molecular characterization of essential components of the *Vibrio anguillarum* plasmid-mediated iron-transport system. J. Biol. Chem. **263:**2853–2860.
3. **Adams, M. D., and D. L. Oxender.** 1989. Bacterial periplasmic binding protein tertiary structures. J. Biol. Chem. **264:**15739–15742.
4. **Adams, M. D., L. M. Wagner, T. J. Graddis, R. Landick, T. K. Antonucci, A. L. Gibson, and D. L. Oxender.** 1990. Nucleotide sequence and genetic characterization reveal six essential genes for the LIV-I and LS transport systems of *Escherichia coli*. J. Biol. Chem. **265:**11436–11443.
5. **Ahlem, C., W. Huisman, G. Nesland, and A. S. Dahms.** 1982. Purification and properties of a periplasmic D-xylose-binding protein from *Escherichia coli*. J. Biol. Chem. **257:**2926–2931.
6. **Akasaka, M., J. Mizoguchi, and K. Takahashi.** 1990. Human cDNA sequence of a novel glutathione peroxidase-related protein. Nucleic Acids Res. **18:**4619.
7. **Alloing, G., M. C. Trombe, and J. P. Claverys.** 1990. The *ami* locus of the gram-positive bacterium *Streptococcus pneumoniae* is similar to binding protein-dependent transport operons of Gram-negative bacteria. Mol. Microbiol. **4:**633–644.
8. **Ames, G. F.-L., and A. K. Joshi.** 1990. Energy coupling in bacterial periplasmic permeases. J. Bacteriol. **172:**4133–4137.
9. **Ames, G. F.-L., and H. LeCar.** 1992. ATP-dependent bacterial transporters and cystic fibrosis: analogy between channels and transporters. FASEB J. **6:**2660–2666.
10. **Angerer, A., S. Gaisser, and V. Braun.** 1990. Nucleotide sequences of the *sfuA*, *sfuB*, and *sfuC* genes of *Serratia marcescens* suggest a periplasmic-binding-protein-dependent iron transport mechanism. J. Bacteriol. **172:**572–578.
11. **Ankenbauer, R. G., and E. W. Nester.** 1990. Sugar-mediated induction of *Agrobacterium tumefaciens* virulence genes: structural specificity and activities of monosaccharides. J. Bacteriol. **172:**6442–6446.
12. **Anraku, Y.** 1968. Transport of sugars and amino acids in bacteria. I. Purification and specificity of the galactose- and leucine-binding proteins. J. Biol. Chem. **243:**3116–3122.
13. **Argast, M., and W. Boos.** 1979. Purification and properties of the *sn*-glycerol 3-phosphate-binding protein of *Escherichia coli*. J. Biol. Chem. **254:**10931–10935.
14. **Bairoch, A.** 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **19:**2241–2245.
15. **Bairoch, A.** 1992. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **20:**2013–2018.
16. **Barash, H., and Y. S. Halpern.** 1975. Purification and properties of glutamate binding protein from the periplasmic space of *Escherichia coli* K-12. Biochim. Biophys. Acta **386:**168–180.
17. **Bell, A. W., S. D. Buckel, J. M. Groarke, J. N. Hope, D. H. Kingsley, and M. A. Hermodson.** 1986. The nucleotide sequence of the *rbsD*, *rbsA* and *rbsC* genes of *Escherichia coli* K12. J. Biol. Chem. **261:**7652–7658.
18. **Berger, E. A., and L. A. Heppel.** 1972. A binding protein involved in the transport of cystine and diaminopimelic acid in *Escherichia coli*. J. Biol. Chem. **247:**7684–7755.
19. **Berish, S. A., D. R. Kapczynski, and S. A. Morse.** 1990. Nucleotide sequence of the *fbp* gene from *Neisseria meningitidis*. Nucleic Acids Res. **18:**4596.
20. **Boos, W.** 1974. Pro and contra transport carriers: the role of the galactose-binding protein in the β-methylgalactoside transport system of *Escherichia coli*. Curr. Top. Membr. Transp. **5:**51–136.
21. **Bourret, R. B., K. A. Borkovich, and M. I. Simon.** 1991. Signal transduction pathways involving protein phosphorylation in prokaryotes. Annu. Rev. Biochem. **60:**401–441.
22. **Bradbeer, C., J. S. Kenley, D. R. Di Masi, and J. Leighton.** 1978. Transport of vitamin B_{12} in *Escherichia coli*. Corrinoid specificities of the periplasmic B_{12}-binding protein and of energy dependent B_{12} transport. J. Biol. Chem. **253:**1347–1352.
23. **Buckenmeyer, G. K., and M. A. Hermodson.** 1983. The amino acid sequence of D-ribose-binding protein from *Salmonella typhimurium* ST1. J. Biol. Chem. **258:**12957.
24. **Burnett, W. V., J. Henner, and T. Eckhardt.** 1987. The nucleotide sequence of the gene coding for XP55, a major secreted protein from *Streptomyces lividans*. Nucleic Acids Res. **15:**3926.
25. **Callahan, H. L., and S. M. Beverley.** 1991. Heavy metal resistance: a new role for P-glycoproteins in *Leishmania*. J. Biol. Chem. **266:**18427–18430.
26. **Cangelosi, G. A., R. G. Ankenbauer, and E. W. Nester.** 1990. Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. Proc. Natl. Acad. Sci. USA **87:**6708–6712.
27. **Cangelosi, G. A., G. Martinetti, J. A. Leigh, C. C. Lee, C. Theines, and E. W. Nester.** 1989. Role of *Agrobacterium tumefaciens* ChvA protein in export of β-1,2-glucan. J. Bacteriol. **171:**1609–1615.
28. **Chen, C.-J., J. E. Chin, K. Ueda, D. P. Clark, I. Pastan, M. M. Gottesman, and I. B. Roninson.** 1986. Internal duplication and homology with bacterial transport proteins in the *mdr1* (P-glycoprotein) gene from multidrug-resistant human cells. Cell **47:**381–389.
29. **Chen, C.-M., T. K. Misra, S. Silver, and B. P. Rosen.** 1986. Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. J. Biol. Chem. **261:**15030–15038.
30. **Chen, C.-M., Q.-Z. Ye, Z. Zhu, B. L. Wanner, and C. T. Walsh.** 1990. Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the *phn* (*psiD*) genes involved in alkylphosphonate uptake and C-P cyclase activity in *Escherichia coli* B. J. Biol. Chem. **265:**4461–4471.
31. **Coulton, J. W., P. Mason, and D. D. Allatt.** 1987. *fhuC* and *fhuD* genes for iron(III)-ferrichrome transport into *Escherichia coli* K-12. J. Bacteriol. **169:**3844–3849.
32. **Cox, G. B., D. Webb, J. Godovac-Zimmermann, and H. Rosenberg.** 1988. Arg-220 of the PstA protein is required for phosphate transport through the phosphate-specific transport system in *Escherichia coli* but not for alkaline phosphatase repression. J. Bacteriol. **170:**2283–2286.
33. **Cox, G. B., D. Webb, and H. Rosenberg.** 1989. Specific amino acid residues in both the PstB and PstC proteins are required for phosphate transport by the *Escherichia coli* Pst system. J. Bacteriol. **171:**1531–1534.
34. **Dahl, M. K., E. Francoz, W. Saurin, W. Boos, M. D. Manson, and M. Hofnung.** 1989. Comparison of sequences from the *malB* regions of *Salmonella typhimurium* and *Enterobacter aerogenes* with *Escherichia coli* K12: a potential new regulatory site in the interoperonic region. Mol. Gen. Genet. **218:**199–207.

34a. **Dassa, E., and M. Hofnung.** Personal communication.

35. **Davidson, A. L., and H. Nikaido.** 1990. Overproduction, solubilization, and reconstitution of the maltose transport system from *Escherichia coli*. J. Biol. Chem. **265:**4254–4260.
36. **Dayhoff, M. O., W. C. Barker, and L. T. Hunt.** 1983. Establishing homologies in protein sequences. Methods Enzymol. **91:**524–545.

37. **Dean, D. A., A. L. Davidson, and H. Nikaido.** 1989. Maltose transport in membrane vesicles of *Escherichia coli* is linked to ATP hydrolysis. Proc. Natl. Acad. Sci. USA **86**:9134–9138.
38. **Dean, D. A., J. D. Fikes, K. Gehring, P. J. Bassford, Jr., and H. Nikaido.** 1989. Active transport of maltose in membrane vesicles obtained from *Escherichia coli* cells producing tethered maltose-binding protein. J. Bacteriol. **171**:503–510.
39. **Dean, D. A., J. Reizer, H. Nikaido, and M. H. Saier, Jr.** 1990. Regulation of the maltose transport system of *Escherichia coli* by the glucose-specific Enzyme III of the phosphoenolpyruvate-sugar phosphotransferase system. Characterization of inducer exclusion-resistant mutants and reconstitution of inducer exclusion in proteoliposomes. J. Biol. Chem. **265**:21005–21010.
40. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analyses for the VAX. Nucleic Acids Res. **12**:387–395.
41. **Doolittle, R. F.** 1981. Similar amino acid sequences: chance or common ancestry? Science **214**:149–159.
42. **Doolittle, R. F.** 1986. Of Urfs and Orfs. University Science Books, Mill Valley, Calif.
43. **Doolittle, R. F., and D.-F. Feng.** 1990. A nearest neighbor procedure for relating progressively aligned amino acid sequences. Methods Enzymol. **183**:659–669.
44. **Dreesen, T. D., D. H. Johnson, and S. Henikoff.** 1988. The brown protein of *Drosophila melanogaster* is similar to the white protein and to components of active-transport complexes. Mol. Cell. Biol. **8**:5206–5215.
45. **Dreyfuss, J.** 1965. Characterization of a sulfate-thiosulfate-transport system in *Salmonella typhimurium*. J. Biol. Chem. **239**:2292–2297.
46. **Dudler, T., C. Schmidhauser, R. W. Parish, R. E. H. Wettenhall, and T. Schmidt.** 1988. A mycoplasma high-affinity transport system and the in vivo invasiveness of mouse sarcoma cells. EMBO J. **7**:3963–3970.
47. **Duplay, P., H. Bedouelle, A. Fowler, I. Zabin, W. Saurin, and M. Hofnung.** 1984. Sequences of the *malE* gene and of its product, the maltose-binding protein of *Escherichia coli* K12. J. Biol. Chem. **259**:10606–10613.
48. **Endicott, J. A., F. Sarangi, and V. Ling.** 1991. Complete CDNA sequences encoding the Chinese hamster P-glycoprotein gene family. DNA Sequence **2**:89–101.
49. **Evans, I. J., and J. A. Downie.** 1986. The *nodI* gene product of *Rhizobium leguminosarum* is closely related to ATP-binding bacterial transport proteins: nucleotide sequence analysis of the *nodI* and *nodJ* genes. Gene **43**:95–101.
49a. **Fagan, M. J., and M. H. Saier, Jr.** J. Mol. Evol., in press.
50. **Felmlee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163**:94–105.
51. **Feng, D.-F., and R. F. Doolittle.** 1990. Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol. **183**:375–387.
52. **Fikes, J. D., and P. J. Bassford, Jr.** 1987. Export of unprocessed precursor maltose-binding proteins to the periplasm of *Escherichia coli* cells. J. Bacteriol. **169**:2352–2359.
53. **Friedrich, M. J., L. C. Deveaux, and R. J. Kadner.** 1986. Nucleotide sequence of the *btuCED* genes involved in vitamin B_{12} transport in *Escherichia coli* and homology with components of periplasmic-binding-protein-dependent transport systems. J. Bacteriol. **167**:928–934.
54. **Frosch, M., U. Edwards, K. Brousset, B. Krausse, and C. Weisgerber.** 1991. Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. Mol. Microbiol. **5**:1251–1263.
55. **Furlong, C. E.** 1987. Osmotic-shock-sensitive transport systems, p. 768–796. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
56. **Furlong, C. E., and L. A. Heppel.** 1971. Leucine binding proteins from *Escherichia coli*. Methods Enzymol. **17B**:639–643.
57. **Furlong, C. E., and J. H. Weiner.** 1970. Purification of a leucine-specific binding protein from *Escherichia coli*. Biochem. Biophys. Res. Commun. **38**:1076–1083.
58. **Furuchi, T., K. Kashiwagi, H. Kobayashi, and K. Igarashi.** 1991. Characteristics of the gene for a spermidine and putrescine transport system that maps at 15 min on the *Escherichia coli* chromosome. J. Biol. Chem. **266**:20928–20933.
59. **Gardina, P., C. Conway, M. Kossman, and M. Manson.** 1992. Aspartate and maltose-binding protein interact with adjacent sites in the tar chemotactic signal transducer of *Escherichia coli*. J. Bacteriol. **174**:1528–1536.
60. **Gartner, J., H. Moser, and D. Vallee.** 1992. Mutations in the 704 peroxisomal membrane protein gene in Zellweger syndrome. Nature Genet. **1**:16–23.
61. **Geistlich, M., R. Losick, J. R. Turner, and R. N. Rao.** 1992. Characterization of a novel regulatory gene governing the expression of a polyketide synthase gene in *Streptomyces ambofaciens*. Mol. Microbiol. **6**:2019–2029.
62. **Gibson, M. M., M. Price, and C. F. Higgins.** 1984. Genetic characterization and molecular cloning of the tripeptide permease (*tpp*) genes of *Salmonella typhimurium*. J. Bacteriol. **160**:122–130.
63. **Gilson, E., G. Alloing, T. Schmidt, J.-P. Claverys, R. Dudler, and M. Hofnung.** 1988. Evidence for high affinity binding-protein dependent transport systems in Gram-positive bacteria and in *Mycoplasma*. EMBO J. **7**:3971–3974.
64. **Gilson, L., H. K. Mahanty, and R. Kolter.** 1990. Genetic analysis of an MDR-like export system: the secretion of colicin V. EMBO J. **9**:3875–3884.
65. **Glaser, P., H. Sakamoto, J. Bellalou, A. Ullmann, and A. Danchin.** 1988. Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*. EMBO J. **7**:3997–4004.
66. **Goodell, E. W., and C. F. Higgins.** 1987. Uptake of cell wall peptides by *Salmonella typhimurium* and *Escherichia coli*. J. Bacteriol. **169**:3861–3865.
67. **Gowrishankar, J.** 1989. Nucleotide sequence of the osmoregulatory *proU* operon of *Escherichia coli*. J. Bacteriol. **171**:1923–1931.
67a. **Groisman, E. A.** Personal communication.
68. **Gros, P., J. Croop, and D. Housman.** 1986. Mammalian multidrug resistance gene: complete cDNA sequence indicating strong homology to bacterial transport proteins. Cell **47**:371–380.
69. **Gros, P., M. Raymond, J. Bell, and D. Housman.** 1988. Cloning and characterization of a second member of the mouse *mdr* gene family. Mol. Cell. Biol. **8**:2770–2778.
70. **Grossman, A. D., and R. Losick.** 1988. Extracellular control of spore formation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **85**:4369–4373.
71. **Guilfoile, P. G., and C. R. Hutchinson.** 1991. A bacterial analog of the *mdr* gene of mammalian tumor cells is present in *Streptomyces peucetius*, the producer of daunorubicin and doxorubicin. Proc. Natl. Acad. Sci. USA **88**:8553–8557.
72. **Guyer, C. A., D. G. Morgan, N. Osheroff, and J. V. Staros.** 1985. Purification and characterization of a periplasmic oligopeptide binding protein from *Escherichia coli*. J. Biol. Chem. **260**:10812–10816.
73. **Gy, G. J., and S. Silver.** 1992. Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. Proc. Natl. Acad. Sci. USA **89**:9474–9478.
74. **Hellinga, H. W., and P. R. Evans.** 1985. Nucleotide sequence and high-level expression of the major *Escherichia coli* phosphofructokinase. Eur. J. Biochem. **149**:363–373.
75. **Higgins, C. F.** 1992. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. **8**:67–113.
76. **Higgins, C. F., and G. F.-L. Ames.** 1981. Two periplasmic transport proteins which interact with a common membrane receptor show extensive homology complete nucleotide sequences. Proc. Natl. Acad. Sci. USA **78**:6038–6042.

77. **Higgins, C. F., P. D. Haag, K. Nikaido, F. Ardeshir, G. Garcia, and G. F.-L. Ames.** 1982. Complete nucleotide sequence and identification of membrane components of the histidine transport operon of *S. typhimurium*. Nature (London) **298**:723–727.
78. **Higgins, C. F., S. C. Hyde, M. M. Mimmack, U. Gileadi, D. R. Gill, and M. P. Gallagher.** 1990. Binding protein-dependent transport systems. J. Bioenerg. Biomembr. **22**:571–592.
79. **Highlander, S. K., M. Chidambarum, M. M. Engler, and G. M. Weinstock.** 1989. DNA sequence of the *Pasteurella haemolytica* leukotoxin gene cluster. DNA **8**:15–28.
80. **Hiles, I. D., M. P. Gallagher, D. J. Jamieson, and C. F. Higgins.** 1987. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. J. Mol. Biol. **195**:125–142.
81. **Hiles, I. D., and C. F. Higgins.** 1986. Peptide uptake by *Salmonella typhimurium*. The periplasmic oligopeptide-binding protein. Eur. J. Biochem. **158**:561–567.
82. **Hoch, J. A.** 1976. Genetics of bacterial sporulation. Adv. Genet. **18**:69–99.
83. **Hoch, J. A.** 1993. The phosphorelay signal transduction pathway in the initiation of *Bacillus subtilis* sporulation. J. Cell. Biochem. **51**:55–61.
84. **Hogg, R. W., and M. A. Hermodson.** 1977. Amino acid sequence of the L-arabinose-binding protein from *Escherichia coli* B/r. J. Biol. Chem. **252**:5135–5141.
85. **Hoshino, T., and K. Kose.** 1990. Cloning, nucleotide sequences, and identification of products of the *Pseudomonas aeruginosa* PAO *bra* genes, which encode the high-affinity branched-chain amino acid transport system. J. Bacteriol. **172**:5531–5539.
86. **Hoshino, T., K. Kose-Terai, and K. Sato.** 1992. Solubilization and reconstitution of the *Pseudomonas aeruginosa* high affinity branched-chain amino acid transport system. J. Biol. Chem. **267**:21313–21318.
87. **Hryniewicz, M., A. Sirko, A. Palucha, A. Böck, and D. Hulanicka.** 1990. Sulfate and thiosulfate transport in *Escherichia coli* K-12: identification of a gene encoding a novel protein involved in thiosulfate binding. J. Bacteriol. **172**:3358–3366.
88. **Huang, M.-L. W., G. A. Cangelosi, W. Halperin, and E. W. Nester.** 1990. A chromosomal *Agrobacterium tumefaciens* gene required for effective plant signal transduction. J. Bacteriol. **172**:1814–1822.
89. **Isihara, H., and R. W. Hogg.** 1980. Amino acid sequence of the sulfate-binding protein from *Salmonella typhimurium* LT2. J. Biol. Chem. **255**:4614–4618.
90. **Iwashima, A., A. Matsuura, and Y. Nose.** 1971. Thiamine-binding protein of *Escherichia coli*. J. Bacteriol. **108**:1419–1421.
91. **Jamieson, D. J., and C. F. Higgins.** 1984. Anaerobic and leucine dependent expression of a peptide transport gene in *Salmonella typhimurium*. J. Bacteriol. **160**:131–136.
92. **Kamijo, K., S. Taketani, S. Yokota, T. Osumi, and T. Hashimoto.** 1989. The 70-kDa peroxisomal membrane protein is a member of the Mdr (P-glycoprotein)-related ATP-binding protein superfamily. J. Biol. Chem. **265**:4534–4540.
93. **Kanehisa, M.** 1982. Los Alamos sequence analysis package for nucleic acids and proteins. Nucleic Acids Res. **10**:183–196.
94. **Kashiwagi, K., Y. Yamaguchi, Y. Sakai, H. Kobayashi, and K. Igarashi.** 1990. Identification of the polyamine-induced protein as a periplasmic oligopeptide binding protein. J. Biol. Chem. **265**:8387–8391.
95. **Kellermann, O., and S. Szmelcman.** 1974. Active transport of maltose in *Escherichia coli* K12. Eur. J. Biochem. **47**:139–149.
96. **Kerppola, R. E., and G. F.-L. Ames.** 1992. Topology of the hydrophobic membrane-bound components of the histidine periplasmic permease. Comparison with other members of the family. J. Biol. Chem. **267**:2329–2336.
97. **Kossmann, M., C. Wolff, and M. D. Manson.** 1988. Maltose chemoreceptor of *Escherichia coli*: interaction of maltose-binding protein and the Tar signal transducer. J. Bacteriol. **170**:4516–4521.
98. **Köster, W. L., L. A. Actis, L. S. Waldbeser, M. E. Tolmasky, and J. H. Crosa.** 1991. Molecular characterization of the iron transport system mediated by the pJM1 plasmid in *Vibrio anguillarum* 775. J. Biol. Chem. **266**:23829–23833.
99. **Kostler, W., and V. Braun.** 1986. Iron hydroxamate transport of *Escherichia coli*: nucleotide sequence of the *fhuB* gene and identification of the protein. Mol. Gen. Genet. **204**:435–442.
100. **Krishnan, M., J. W. Burgner, W. S. Chilton, and S. B. Gelvin.** 1991. Transport of nonmetabolizable opines by *Agrobacterium tumefaciens*. J. Bacteriol. **173**:903–905.
101. **Kroll, J. S., I. Hopkins, and E. R. Moxon.** 1988. Capsule loss in *H. influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. Cell **53**:347–356.
102. **Kroll, J. S., B. Loynds, L. N. Brophy, and E. R. Moxon.** 1990. The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. Mol. Microbiol. **4**:1853–1862.
103. **Kuchler, K., R. E. Sterne, and J. Thorner.** 1989. *Saccharomyces cerevisiae STE6* gene product: a novel pathway for protein export in eukaryotic cells. EMBO J. **8**:3973–3984.
104. **Kustu, S. G., N. C. McFarland, S. P. Hui, B. Esmon, and G. F. Ames.** 1979. Nitrogen control in *Salmonella typhimurium*: coregulation of synthesis of glutamine synthetase and amino acid transport systems. J. Bacteriol. **138**:218–234.
105. **Lacks, S. A., J. J. Dunn, and B. Greenberg.** 1982. Identification of base mismatches recognized by the heteroduplex-DNA-repair system of *Streptococcus pneumoniae*. Cell **31**:327–336.
106. **Landick, R., and D. L. Oxender.** 1985. The complete nucleotide sequences of the *Escherichia coli* LIV-BP and LS-BP genes. Implications for the mechanism of high-affinity branched-chain amino acid transport. J. Biol. Chem. **260**:8257–8261.
107. **Laudenbach, D. E., and A. R. Grossman.** 1991. Characterization and mutagenesis of sulfur-regulated genes in a cyanobacterium: evidence for function in sulfate transport. J. Bacteriol. **173**:2739–2750.
108. **Lavitola, A., M. Vanni, P. M. V. Martin, and C. B. Bruni.** 1992. Cloning and characterization of a *Neisseria* gene homologous to *hisJ* and *argT* of *Escherichia coli* and *Salmonella typhimurium*. Res. Microbiol. **143**:295–305.
109. **Leive, L., and B. D. Davis.** 1965. The transport of diaminopimelate and cystine in *Escherichia coli*. J. Biol. Chem. **240**:4362–4369.
110. **Lengeler, J. W., J. Bockmann, H. Heuel, and F. Titgemeyer.** 1992. The enzymes II of the PTS as carbohydrate transport systems: what the evolutionary studies tell us on their structure and function, p. 77–85. *In* E. Quagliariello and F. Palmieri (ed.), Molecular mechanisms of transport. Elsevier Science Publishing, Inc., New York.
111. **Letoffe, S., P. Delepelaire, and C. Wandersman.** 1990. Protease secretion by *Erwinia chrysanthemi*: the specific secretion functions are analogous to those of *Escherichia coli* α-haemolysin. EMBO J. **9**:1375–1382.
112. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
113. **Lukat, G. S., and J. B. Stock.** 1993. Response regulation in bacterial chemotaxis. J. Cell. Biochem. **51**:41–46.
114. **Magota, K., N. Otsuji, T. Miki, T. Horiuchi, S. Tsunasawa, J. Kondo, F. Sakiyama, M. Amemura, T. Morita, H. Shinagawa, and A. Nakata.** 1984. Nucleotide sequence of the *phoS* gene, the structural gene for the phosphate-binding protein of *Escherichia coli*. J. Bacteriol. **157**:909–917.
115. **Mahoney, W. C., R. W. Hogg, and M. A. Hermodson.** 1981. The amino acid sequence of the D-galactose-binding protein from *Escherichia coli* B/r. J. Biol. Chem. **256**:4350–4356.
116. **Makino, K., S.-K. Kim, H. Shinagawa, M. Amemura, and A. Nakata.** 1991. Molecular analysis of the cryptic and functional *phn* operons for phosphonate use in *Escherichia coli* K-12. J. Bacteriol. **173**:2665–2672.
117. **Makino, K., H. Shinagawa, M. Amemura, T. Kawamoto, M. Yamada, and A. Nakata.** 1989. Signal transduction in the phosphate regulon of *Escherichia coli* involves phosphotransfer between PhoR and PhoB proteins. J. Mol. Biol. **210**:551–559.
118. **Makino, K., H. Shinagawa, M. Amemura, S. Kimura, A.**

Nakata, and A. Ishihama. 1988. Regulation of the phosphate regulon of *Escherichia coli*: activation of *pstS* transcription by PhoB protein *in vitro*. J. Mol. Biol. **203:**85–95.

119. **Marger, M. D., and M. H. Saier, Jr.** 1993. A major superfamily of transmembrane facilitators catalyzing uniport, symport and antiport. Trends Biochem. Sci. **18:**13–20.
120. **Martineau, P., W. Saurin, M. Hofnung, J. C. Spurlino, and F. A. Quiocho.** 1990. Progress in the identification of interaction sites on the periplasmic maltose binding protein from *E. coli*. Biochimie **72:**397–402.
121. **Mathiopoulos, C., J. P. Mueller, F. J. Slack, C. G. Murphy, S. Patankar, G. Bukusoglu, and A. L. Sonenshein.** 1991. A *Bacillus subtilis* dipeptide transport system expressed early during sporulation. Mol. Microbiol. **5:**1903–1913.
122. **Mauzy, C. A., and M. A. Hermodson.** 1992. Structural and functional analyses of the repressor, RbsR, of the ribose operon of *Escherichia coli*. Protein Sci. **1:**831–842.
123. **Mauzy, C. A., and M. A. Hermodson.** 1992. Structural homology between *rbs* repressor and ribose binding protein implies functional similarity. Protein Sci. **1:**843–849.
124. **May, G., E. Faatz, J. M. Lucht, M. Haardt, M. Bolliger, and E. Bremer.** 1989. Characterization of the osmoregulated *Escherichia coli proU* promoter and identification of ProU as a membrane-associated protein. Mol. Microbiol. **3:**1521–1531.
125. **McGrath, J. P., and A. Varshavsky.** 1989. The yeast *STE6* gene encodes a homologue of the mammalian multidrug resistance P-glycoprotein. Nature (London) **340:**400–404.
126. **Medveczky, N., and H. Rosenberg.** 1970. The phosphate-binding protein of *Escherichia coli*. Biochim. Biophys. Acta **211:**158–168.
127. **Metcalf, W. W., and B. L. Wanner.** 1991. Involvement of the *Escherichia coli phn* (*psiD*) gene cluster in assimilation of phosphorus in the form of phosphonates, phosphite, P_i esters, and P_i. J. Bacteriol. **173:**587–600.
128. **Mietzner, T. A., G. Bolan, G. K. Schoolnik, and S. A. Morese.** 1987. Purification and characterization of the major iron-regulated protein expressed by pathogenic Neisseria. J. Exp. Med. **165:**1041–1057.
129. **Miller, D. M., III, J. S. Olson, J. W. Pflugrath, and F. A. Quiocho.** 1983. Rates of ligand binding to periplasmic proteins involved in bacterial transport and chemotaxis. J. Biol. Chem. **258:**13665–13672.
130. **Mowbray, S. L., R. D. Smith, and L. B. Cole.** 1990. Structure of the periplasmic glucose/galactose receptor of *Salmonella typhimurium*. Receptor **1:**41–53.
131. **Müller-Hill, B.** 1983. Sequence homology between Lac and Gal repressors and three sugar-binding periplasmic proteins. Nature (London) **302:**163–164.
132. **Newcomer, M. E., G. L. Gilliland, and F. A. Quiocho.** 1981. L-Arabinose-binding protein-sugar complex at 2.4 Å resolution. J. Biol. Chem. **256:**13213–13217.
133. **Newcomer, M. E., B. A. Lewis, and F. A. Quiocho.** 1981. The radius of gyration of L-arabinose-binding protein decreases upon binding of ligand. J. Biol. Chem. **256:**13218–13222.
134. **Nielsen, J. B. K., M. P. Caulfield, and J. O. Lampen.** 1981. Lipoprotein nature of *Bacillus licheniformis* membrane penicillinase. Proc. Natl. Acad. Sci. USA **78:**3511–3515.
135. **Nielsen, J. B. K., and J. O. Lampen.** 1982. Glyceride-cysteine lipoproteins and secretion by gram-positive bacteria. J. Bacteriol. **152:**315–322.
136. **Nikaido, H., and M. H. Saier, Jr.** 1992. Transport proteins in bacteria: common themes in their design. Science **258:**936–942.
137. **Nishimune, T., and R. Hayashi.** 1971. Thiamine-binding protein and thiamine uptake by *Escherichia coli*. Biochim. Biophys. Acta **224:**573–583.
138. **Nohno, T., T. Saito, and J. Hong.** 1986. Cloning and complete nucleotide sequence of the *Escherichia coli* glutamine permease operon (*glnHPQ*). Mol. Gen. Genet. **205:**260–269.
139. **Nucifora, G., L. Chu, S. Silver, and T. K. Misra.** 1989. Mercury operon regulation by the *merR* gene of the organomercurial resistance system of plasmid pDU1358. J. Bacteriol. **171:**4241–4247.
140. **O'Hare, K., C. Murphy, R. Levis, and G. M. Rubin.** 1984. DNA sequence of the white locus of *Drosophila melanogaster*. J. Mol. Biol. **180:**437–455.
141. **Olson, E. R., D. S. Dunyak, L. M. Jurss, and R. A. Poorman.** 1991. Identification and characterization of *dppA*, an *Escherichia coli* gene encoding a periplasmic dipeptide transport protein. J. Bacteriol. **173:**234–244.
142. **Oshima, R. G., R. C. Willis, C. E. Furlong, and J. A. Schneider.** 1974. Binding assays for amino acids: the utilization of a cystine binding protein from *Escherichia coli* for the determination of acid-soluble cystine in small physiological samples. J. Biol. Chem. **249:**6033–6039.
143. **Ouellette, M., F. Fase-Fowler, and P. Borst.** 1990. The amplified H circle of methotrexate resistant *Leishmania tarentolae* contains a novel P-glycoprotein gene. EMBO J. **9:**1027–1033.
144. **Overdier, G., E. R. Olson, B. D. Erickson, M. M. Ederer, and L. N. Csonka.** 1989. Nucleotide sequence of the transcriptional control region of the osmotically regulated *proU* operon of *Salmonella typhimurium* and identification of the 5′ endpoint of the *proU* mRNA. J. Bacteriol. **171:**4694–4706.
145. **Overduin, P., W. Boos, and J. Tommassen.** 1988. Nucleotide sequence of the *ugp* genes of *Escherichia coli* K-12: homology to the maltose system. Mol. Microbiol. **2:**767–775.
146. **Pardee, A. B.** 1966. Purification and properties of a sulfate-binding protein from *Salmonella typhimurium*. J. Biol. Chem. **241:**5886–5892.
147. **Pavelka, M. S., L. F. Wright, and R. P. Silver.** 1991. Identification of two genes, *kpsM* and *kpsT*, in region 3 of the polysialic acid gene cluster of *Escherichia coli* K1. J. Bacteriol. **173:**4603–4610.
148. **Payne, J. W., and C. Gilvarg.** 1968. Size restriction on peptide utilization in *Escherichia coli*. J. Biol. Chem. **243:**6391–6399.
149. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
150. **Penrose, W. R., G. E. Nichoalds, J. R. Piperno, and D. L. Oxender.** 1968. Purification and properties of a leucine-binding protein from *Escherichia coli*. J. Biol. Chem. **243:**5921–5928.
151. **Perego, M., C. F. Higgins, S. R. Pearce, M. P. Gallagher, and J. A. Hoch.** 1991. The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation. Mol. Microbiol. **5:**173–185.
152. **Petronilli, V., and G. F.-L. Ames.** 1991. Binding protein-independent histidine permease mutants. Uncoupling of ATP hydrolysis from transmembrane signaling. J. Biol. Chem. **266:**16293–16296.
153. **Powis, S. H., I. Mockridge, A. Kelly, L.-A. Kerr, R. Glynne, U. Gileadi, S. Beck, and J. Trowsdale.** 1992. Polymorphism in a second ABC transporter gene located with the class 11 region of the MHC. Proc. Natl. Acad. Sci. USA **89:**1463–1467.
154. **Quiocho, F. A.** 1990. Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. Philos. Trans. R. Soc. Lond. Ser. B **326:**341–351.
155. **Randall, L. L., and S. J. S. Hardy.** 1989. Unity in function in the absence of consensus in sequence: role of leader peptides in export. Science **243:**1156–1159.
156. **Reizer, A., G. M. Pao, and M. H. Saier, Jr.** 1991. Evolutionary relationships among the permease proteins of the bacterial phosphoenolpyruvate: sugar phosphotransferase system. Construction of phylogenetic trees and possible relatedness to proteins of eukaryotic mitochondria. J. Mol. Evol. **33:**179–193.
157. **Reizer, J., K. Finley, D. Kakuda, C. L. MacLeod, A. Reizer, and M. H. Saier, Jr.** 1993. Mammalian integral membrane receptors are homologous to facilitators and antiporters of yeast, fungi and eubacteria. Protein Sci. **2:**20–30.
158. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1990. The cellobiose permease of *Escherichia coli* consists of three proteins and is homologous to the lactose permease of *Staphylococcus aureus*. Res. Microbiol. **141:**1061–1067.
159. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1992. A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. Protein Sci. **1:**1326–1332.
160. **Reizer, J., A. Reizer, M. H. Saier, Jr., and G. R. Jacobson.** 1992. A proposed link between nitrogen and carbon metabo-

lism involving protein phosphorylation in bacteria. Protein Sci. **1:**722–726.

161. **Riordan, J. R., J. M. Rommens, B.-S. Kerem, N. Alon, R. Rozmahel, J. Grzelczak, J. Zielenski, S. Lok, N. Plavsk, J. L. Chou, M. L. Drumm, F. S. Ianuzzi, F. S. Collins, and L. C. Tsui.** 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science **245:** 1066–1073.
162. **Rosen, B. P.** 1971. Basic amino acid transport in *Escherichia coli*. J. Biol. Chem. **246:**3653–3662.
163. **Rosenberg, H., R. G. Gerdes, and K. Chegwidden.** 1977. Two systems for the uptake of phosphate in *Escherichia coli*. J. Bacteriol. **131:**505–511.
164. **Ross, J. I., E. A. Eady, J. H. Cove, W. J. Cunliffe, S. Baumberg, and J. C. Wootton.** 1990. Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super-gene family. Mol. Microbiol. **4:**1207–1214.
165. **Rosteck, P. R., P. A. Reynolds, and C. L. Hershberger.** 1991. Homology between proteins controlling *Streptomyces fradiae* tylosin resistance and ATP-binding transport. Gene **102:**27–32.
166. **Rotman, B., and J. Radojkovic.** 1964. Galactose transport in *Escherichia coli*. J. Biol. Chem. **239:**3153–3156.
167. **Rudner, D. Z., J. R. Le Deaux, K. Ireton, and A. D. Grossman.** 1991. The *spo0K* locus of *Bacillus subtilis* is homologous to the oligopeptide permease locus and is required for sporulation and competence. J. Bacteriol. **173:**1388–1398.
168. **Russell, R. R. B., J. Aduse-Opoku, I. C. Sutcliffe, L. Tao, and J. J. Ferretti.** 1992. A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. J. Biol. Chem. **267:**4631–4637.
169. **Saier, M. H., Jr.** 1993. Protein phosphorylation and signal transduction in bacteria: an introduction. J. Cell. Biochem. **51:**61–68.
170. **Saier, M. H., Jr., and G. R. Jacobson.** 1984. The molecular basis of sex and differentiation: a comparative study of evolution, mechanism and control in microorganisms. Springer-Verlag, New York.
171. **Saier, M. H., Jr., and J. Reizer.** 1991. Families and superfamilies of transport proteins common to prokaryotes and eukaryotes. Curr. Opin. Struct. Biol. **1:**362–368.
172. **Saier, M. H., Jr., P. K. Werner, and M. Müller.** 1989. Insertion of proteins into bacterial membranes: mechanism, characteristics, and comparisons with the eucaryotic process. Microbiol. Rev. **53:**333–366.
173. **Sautereau, A. M., and M. C. Trombe.** 1986. Electric transmembrane potential mutation and resistance to the cationic and amphiphilic antitumoral drugs derived from pyridocarbazole, 2-*N*-methylellipticinium and 2-*N*-methyl-9-hydroxyellipticinium, in *Streptococcus pneumoniae*. J. Gen. Microbiol. **132:** 2637–2641.
174. **Schatz, P. J., and J. Beckwith.** 1990. Genetic analysis of protein export in *Escherichia coli*. Annu. Rev. Genet. **24:**215–248.
175. **Schoner, B., M. Geistlich, P. Rosteck, Jr., R. N. Rao, E. Seno, P. Reynolds, K. Cox, S. Burgett, and C. Hershberger.** 1992. Sequence similarity between macrolide-resistance determinants and ATP-binding transport proteins. Gene **115:**93–96.
176. **Schweizer, H., M. Argast, and W. Boos.** 1982. Characteristics of a binding protein-dependent transport system for *sn*-glycerol-3-phosphate in *Escherichia coli* that is part of the *pho* regulon. J. Bacteriol. **150:**1154–1163.
177. **Shamanna, D. K., and K. E. Sanderson.** 1979. Uptake and catabolism of D-xylose in *Salmonella typhimurium* LT2. J. Bacteriol. **139:**64–70.
178. **Shaw, J. G., M. J. Hamblin, and D. J. Kelly.** 1991. Purification, characterization and nucleotide sequence of the periplasmic C4-dicarboxylate-binding protein (DctP) from *Rhodobacter capsulatus*. Mol. Microbiol. **5:**3055–3062.
179. **Shea, C. M., and M. A. McIntosh.** 1991. Nucleotide sequence and genetic organization of the ferric enterobactin transport system: homology to other periplasmic binding protein-dependent systems in *Escherichia coli*. Mol. Microbiol. **5:**1415–1428.
180. **Shimkets, L. J., and D. Kaiser.** 1982. Murein components rescue developmental sporulation of *Myxococcus xanthus*. J. Bacteriol. **152:**462–470.
181. **Shimoda, N., A. Toyoda-Yamamoto, J. Nagamine, S. Usami, M. Katayama, Y. Sakagami, and Y. Machida.** 1990. Control of expression of *Agrobacterium vir* genes by synergistic actions of phenolic signal molecules and monosaccharides. Proc. Natl. Acad. Sci. USA **87:**6684–6688.

181a. **Shuman, H.** Personal communication.

182. **Sicard, A. M.** 1964. A new synthetic medium for *Diplococcus pneumoniae* and its use for the study of reciprocal transformation at the *amiA* locus. Genetics **50:**31–44.
183. **Sicard, A. M., and H. Ephrussi-Taylor.** 1965. Genetic recombination in DNA-induced transformation of *Pneumococcus*. II. Mapping the *amiA* region. Genetics **52:**1207–1227.
184. **Slack, F. J., J. M. Mueller, M. A. Strauch, C. Mathiopoulos, and A. L. Sonenshein.** 1991. Transcriptional regulation of a *Bacillus subtilis* dipeptide transport operon. Mol. Microbiol. **5:**1915–1925.
185. **Smith, A. N., G. J. Boulnois, and I. S. Roberts.** 1990. Molecular analysis of the *Escherichia coli* K5 *kps* locus. Mol. Microbiol. **4:**1863–1869.
186. **Smith, D. W.** 1988. A complete and integrated system for DNA protein sequence analysis using VAX/VMS computers. CABIOS **4:**212.
187. **Spurlino, J. C., G.-Y. Lu, and F. A. Quiocho.** 1991. The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J. Biol. Chem. **266:**5202–5219.
188. **Stanfield, S. W., L. Ielpi, D. O'Brochta, D. R. Helinski, and G. S. Ditta.** 1988. The *ndvA* gene product of *Rhizobium meliloti* is required for β-(1→2)glucan production and has homology to the ATP-binding export protein HlyB. J. Bacteriol. **170:**3523–3530.
189. **Staudenmaier, H., B. Van Hove, Z. Yaraghi, and V. Braun.** 1989. Nucleotide sequences of the *fecBCDE* genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. J. Bacteriol. **171:**2626–2633.
190. **Stirling, D. A., C. S. Hulton, L. Waddell, S. F. Park, G. S. Stewart, I. R. Booth, and C. F. Higgins.** 1989. Molecular characterization of the *proU* locus of *Salmonella typhimurium* and *Escherichia coli* encoding osmoregulated glycine betaine transport systems. Mol. Microbiol. **3:**1025–1038.
191. **Stock, J. B., G. S. Lukat, and A. M. Stock.** 1991. Bacterial chemotaxis and the molecular logic of intracellular signal transduction networks. Annu. Rev. Biophys. Biophys. Chem. **20:**109–136.
192. **Stock, J. B., A. J. Ninfa, and A. M. Stock.** 1989. Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol. Rev. **53:**450–490.
193. **Strathdee, C. A., and R. Y. C. Lo.** 1989. Cloning, nucleotide sequence, and characterization of genes encoding the secretion function of the *Pasteurella haemolytica* leukotoxin determinant. J. Bacteriol. **171:**916–928.
194. **Surin, B. P., D. A. Jans, A. L. Fimmel, D. C. Shaw, G. B. Cox, and H. Rosenberg.** 1984. Structural gene for the phosphate-repressible phosphate-binding protein of *Escherichia coli* has its own promoter: complete nucleotide sequence of the *phoS* gene. J. Bacteriol. **157:**772–778.
195. **Surin, B. P., H. Rosenberg, and G. B. Cox.** 1985. Phosphate-specific transport system of *Escherichia coli*: nucleotide sequence and gene-polypeptide relationships. J. Bacteriol. **161:** 189–198.
196. **Sweet, G. D., J. M. Somers, and W. W. Kay.** 1979. Purification and properties of a citrate-binding transport component, the C protein of *Salmonella typhimurium*. Can. J. Biochem. **57:**710–715.
197. **Szmelcman, S., M. Schwartz, T. J. Silhavy, and W. Boos.** 1965. Maltose transport in *Escherichia coli* K12. A comparison of transport kinetics in wild-type and λ-resistant mutants with dissociation constants of the maltose binding protein as measured by fluorescence quenching. Eur. J. Biochem. **65:**13–19.
198. **Tata, F., P. Stanier, C. Wicking, S. Halford, H. Kruyer, N. J.**

Lench, P. J. Scambler, C. Hansen, J. C. Braman, R. Williamson, and B. J. Wainwright. 1991. Cloning the mouse homolog of the human cystic fibrosis transmembrane conductance regulator gene. Genomics **10**:301–307.

199. **Taylor, R. T., S. A. Norrell, and M. L. Hanna.** 1972. Uptake of cyanocobalamin by *Escherichia coli* B: some characteristics and evidence for a binding protein. Arch. Biochem. Biophys. **148**:366–381.
200. **Thompson, J., and J. A. Donkersloot.** 1992. Biosynthesis and distribution of N-carboxyalkyl amino acids (opines) in bacteria. Res. Microbiol. **143**:127–131.
201. **Trandinh, C. C., G. M. Pao, and M. H. Saier, Jr.** 1992. Structural and evolutionary relationships among the immunophilins: two ubiquitous families of ATP-independent chaperonins. FASEB J. **6**:3410–3420.
202. **Treptow, N. A., and H. A. Shuman.** 1988. Allele-specific *malE* mutations that restore interactions between maltose-binding protein and the inner-membrane components of the maltose transport system. J. Mol. Biol. **202**:809–822.
203. **Trombe, M.-C., G. Lanéelle, and A. M. Sicard.** 1984. Characterization of a *Streptococcus pneumoniae* mutant with altered electric transmembrane potential. J. Bacteriol. **158**:1109–1114.
204. **Trowsdale, J., I. Hanson, I. Mockridge, S. Beck, A. Townsend, and A. Kelly.** 1990. Sequences encoded in the class II region of the MHC related to the 'ABC' superfamily of transporters. Nature (London) **348**:741–744.
205. **Tucker, S. J., D. Tannahill, and C. F. Higgins.** Identification and developmental expression of the *Xenopus laevis* cystic fibrosis transmembrane conductance regulator gene. Hum. Mol. Genet., in press.
206. **Valdivia, R. H., L. Wang, and S. C. Winans.** 1991. Characterization of a putative periplasmic transport system for octopine accumulation encoded by *Agrobacterium tumefaciens* Ti plasmid pTiA6. J. Bacteriol. **173**:6398–6405.
207. **Vartak, N. B., J. Reizer, A. Reizer, J. T. Gripp, E. A. Groisman, L.-F. Wu, J. M. Tomich, and M. H. Saier, Jr.** 1991. Sequence and evolution of the FruR protein of *Salmonella typhimurium*: a pleiotropic transcriptional regulatory protein possessing both activator and repressor functions which is homologous to the periplasmic ribose binding protein. Res. Microbiol. **142**:951–963.
208. **Vyas, N. K., M. N. Vyas, and F. A. Quiocho.** 1987. A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. Nature (London) **327**:635–638.
209. **Vyas, N. K., M. N. Vyas, and F. A. Quiocho.** 1991. Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and functional similarity. J. Biol. Chem. **266**:5226–5237.
210. **Wanner, B. L.** 1993. Gene regulation by phosphate in enteric bacteria. J. Cell. Biochem. **51**:47–54.
211. **Wanner, B. L., and W. W. Metcalf.** Molecular genetic studies of a 10.9-kbp operon in *Escherichia coli* for phosphonate uptake and biodegradation. FEMS Microbiol. Lett., in press.
212. **Weickert, M. J., and S. Adhya.** 1992. A family of bacterial regulators homologous to Gal and Lac repressors. J. Biol. Chem. **267**:15869–15874.
213. **Weiner, J. H., C. E. Furlong, and L. A. Heppel.** 1971. A binding protein for L-glutamine and its relation to active transport in *E. coli*. Arch. Biochem. Biophys. **142**:715–717.
214. **Weiner, J. H., and L. A. Heppel.** 1971. A binding protein for glutamine and its relation to active transport in *E. coli*. J. Biol. Chem. **246**:6933–6941.
215. **Widenhorn, K. A., J. M. Somers, and W. W. Kay.** 1989. Genetic regulation of the tricarboxylate transport operon (*tctI*) of *Salmonella typhimurium*. J. Bacteriol. **171**:4436–4441.
216. **Wiech, H., P. Klappa, and R. Zimmerman.** 1991. Protein export in prokaryotes and eukaryotes. Theme with variations. FEBS Lett. **285**:182–188.
217. **Willis, R. C., and C. E. Furlong.** 1974. Purification and properties of a ribose-binding protein from *Escherichia coli*. J. Biol. Chem. **249**:6926–6929.
218. **Willis, R. C., and C. E. Furlong.** 1975. Purification and properties of a periplasmic glutamate-aspartate binding protein from *Escherichia coli* strain W3092. J. Biol. Chem. **250**:2574–2580.
219. **Wilson, S., and R. Drew.** 1991. Cloning and DNA sequence of *amiC*, a new gene regulating expression of the *Pseudomonas aeruginosa* aliphatic amidase, and purification of the *amiC* product. J. Bacteriol. **173**:4914–4921.
220. **Winans, S. C.** 1992. Two-way chemical signaling in *Agrobacterium*-plant interactions. Microbiol. Rev. **56**:12–31.
221. **Wu, C.-T., M. Budding, M. S. Griffin, and J. M. Croop.** 1991. Isolation and characterization of *Drosophila* multidrug resistance gene homologs. Mol. Cell. Biol. **11**:3940–3948.
222. **Wu, H. C.** 1987. Posttranslational modification and processing of membrane proteins in bacteria, p. 37–71. *In* M. Inouye (ed.), Bacterial outer membranes as model systems. Wiley Interscience, New York.
223. **Wu, L., and N. E. Welker.** 1991. Cloning and characterization of a glutamine transport operon of *Bacillus stearothermophilus* NUB36: effect of temperature on regulation of transcription. J. Bacteriol. **173**:4877–4888.
224. **Yamaguchi, K., F. Yu, and M. Inouye.** 1988. A single amino acid determinant of the membrane localization of lipoproteins in *E. coli*. Cell **53**:423–432.
225. **Yoshifuji, T., W. K. Lemna, C. F. Ballard, C. L. Rosenbloom, R. Rozmahel, N. Plavsic, L.-C. Tsui, and A. L. Beaudet.** 1991. Molecular cloning and sequence analysis of the murine cDNA for the cystic fibrosis transmembrane conductance regulator. Genomics **10**:547–550.
226. **Zanker, H., J. von Lintig, and J. Schröder.** 1992. Opine transport genes in the octopine (*occ*) and nopaline (*noc*) catabolic regions in Ti plasmids of *Agrobacterium tumefaciens*. J. Bacteriol. **174**:841–849.
227. **Zumft, W. G., A. Viebrock-Sambale, and C. Braun.** 1990. Nitrous oxide reductase from denitrifying *Pseudomonas stutzeri*. Eur. J. Biochem. **192**:591–599.